



Norwegian School of Economics

Bergen/Dar es Salaam, fall semester, 2014



How the Lab meets the Field in the midst of Young Entrepreneurs

*A study of how experimental evidence on entrepreneurial traits
materialise in real life*

Una Cecilie Friis og Hild Eirin Strømme

Supervisor: Kjetil Bjorvatn

Master thesis in Economics and Business Administration

Major in Energy, Natural Resources and the Environment (ENE)

& Business Analysis and Performance Management (BUS)

NORWEGIAN SCHOOL OF ECONOMICS

This thesis was written as a part of the Master of Science in Economics and Business Administration at NHH. Please note that neither the institution nor the examiners are responsible – through the approval of this thesis – for the theories and methods used, or results and conclusions drawn in this work.

Abstract

Researchers finding answers to economic questions have in the last decades increasingly been using experimental methods to do so. The link between the laboratory evidence and how this is extrapolated to the outside world is, however, still an unexplored area in literature. Many lab experiments have been dedicated to analysing people's inclination to compete in a laboratory setting, but few have explored how this and other entrepreneurial traits materialise in the field. In a setting of young entrepreneurs establishing their businesses in urban Tanzania, we analyse how their laboratory profile corresponds to field outcomes and choices. The lab evidence is on business knowledge and skills, time preferences, competitiveness, willingness to take risk, self-confidence, and finally, ambitions to become an entrepreneur. The evidence on field outcomes is on business survival, sales and profits, while the evidence on business choices is on investment, business purchases and savings. We expand the quantitative analysis by telling the story of the participants, and the contextual constraints that they face.

We find a strong association between competitiveness and sales, and to some degree also profits, and with investment choices in the field. We find less consistent association with some of the other lab evidence. Our findings thus suggest that *competitiveness* is a key entrepreneurial trait to shape outcomes and choices. We find knowledge to be the most important predictor of higher savings.

Acknowledgements

This master thesis finalizes our degrees in respectively *Energy, Natural Resources and the Environment* (ENE) and in *Business Analysis and Performance Management* at Norwegian School of Economics (NHH), and also meets the last requirements for our CEMS Master's in International Management degree.

We would like to express our gratitude to Professor Kjetil Bjorvatn for his support and guidance throughout the process of writing this master thesis. His enthusiasm and help when we encountered problems in the process was highly appreciated. We are also very grateful for the opportunity to take part in the exciting Social Lab project in Tanzania, and for his ideas on the topic and angle for our analysis.

Clearly, this study would not have been possible without the hard work of Jacqueline Mgumia and the rest of her research team, and the pool of participants in the Social Lab Project in Tanzania. We would like to thank Mgumia herself, and Julieth Kweka, Chambi Chachage and Sudi who helped us greatly in understanding the culture, the dataset and the research on the area, and in our meetings with some of the survey participants.

We raise our thanks to NHH for all the opportunities the school has provided us in regards to our education and in the form of exchange semesters and internships. We especially thank The Choice Lab for the opportunity of participating in research fieldwork in Tanzania, and to our fellow research assistants there. Furthermore, we would like to thank all the friends we have made throughout the years here at NHH.

Finally, we would like to thank our families and friends for their support and patience throughout our studies, and the writing process of this thesis.

Thank you! Tusen takk! Asante sana!

Norwegian School of Economics
Bergen, Norway. 22. December 2014


Una Cecilie Friis

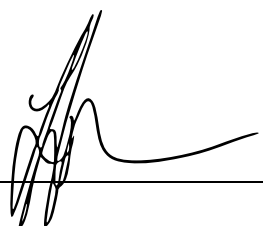

Hild Eirin Strømme

Table of Contents

Abstract.....	2
Acknowledgements	3
Contents	4
List of tables	7
List of abbreviations	7
1. Introduction.....	9
Outline	11
2. Literature review	12
2.1 Why look at entrepreneurs?	12
2.2 The lab – field relationship	12
2.3 Entrepreneurship	14
2.3.1 Competitiveness and gender differences.....	15
2.4 The Context. Youth and Entrepreneurship in Tanzania	16
2.4.1 Education in Tanzania.....	18
Entrepreneurship in urban areas	19
3. Methodology	21
3.1 The Lab Data	22
3.1.1 The Ruka Juu Project	22
3.1.2 The general and socioeconomic background	23
3.1.3 The entrepreneurship indicators.....	24
Knowledge.....	25
Entrepreneurial mind-set	26
Ambitions to become an entrepreneur.....	29
3.1.4 Comparing our pool of participants to the RJ dataset.....	30
Background variables	31
Lab indices.....	31
3.2 The Field Data	33
3.2.1 The Social Lab Project.....	33
Interviews and field visits.....	34

3.2.2	Data cleansing and reassuring data quality.....	35
3.2.3	Key field variables.....	36
	Measuring business outcomes and entrepreneurial choices	36
	Aggregating data.....	38
	Business survival	38
	Business sales.....	38
	Profits.....	38
	Investments and business purchases.....	39
	Savings	39
3.2.4	Description of the field performance	39
	Business sectors.....	39
	Business characteristics	40
	Performance	40
	Weekly development over the year	42
4.	Results	45
4.1	Correlation analysis	45
4.1.1	Spearman correlation test.....	45
4.1.2	Business outcomes	46
	Business survival	47
	Sales	47
	Profits.....	48
	Summary of business outcomes	48
4.1.3	Entrepreneurial choices	49
	Investments	49
	Business Purchases	50
	Savings	50
	Summary of entrepreneurial choices	51
	Main findings from the correlation analysis	51
4.1.4	Constraints and business survival	52
4.2	Regression analysis.....	54
4.2.1	Business outcomes	55
	Sales	55
	Profits.....	56
	Summary of business outcomes	56
4.2.2	Entrepreneurial choices	57
	Investments	57

Business purchases.....	58
Savings.....	58
Summary of entrepreneurial choices.....	59
4.2.3 Findings from the regression analysis	59
4.3 Gender differences	60
4.4 Limitations and validity	62
4.5 Other analytical procedures considered	63
5. Summary and concluding remarks	65
References.....	66
Appendix	69
Table of contents - Appendix	69
List of tables - Appendix.....	69
A. Normality test	70
A.1 Q-plots for normality checking.....	71
B. Financial Diary questionnaire.....	72
C. List of participants, business and lab performance	76
D. Weekly business development over the year.....	78
E. Supplementary correlation tests on competitive.....	79
E.1 Partial correlation on competitive	79
E.2 Point biserial correlation test on competitive	80
F. Regression results	81
References - Appendix	83

List of tables

<i>Table 3-1: Background summary statistics from RJ 2011</i>	24
<i>Table 3-2: Knowledge summary statistics</i>	26
<i>Table 3-3: Mind-set summary statistics</i>	29
<i>Table 3-4: Ambitions summary statistics</i>	30
<i>Table 3-5: Background summary statistics from RJ 2011</i>	31
<i>Table 3-6: Lab indices summary statistics from RJ 2011: A presentation of the difference in the mean values between our pool of participants and the other two groups.</i>	32
<i>Table 3-7: Key field variables summary statistics</i>	34
<i>Table 3-8: Business sectors</i>	39
<i>Table 3-9: Growth per week in TZS</i>	44
<i>Table 4-1: Business outcomes correlation matrix</i>	46
<i>Table 4-2: Entrepreneurial choices correlation matrix</i>	49
<i>Table 4-3: Reasons why 20 participants dropped out of the SL Project</i>	52
<i>Table 4-4: Regression results of business outcomes on indices. 29 participants</i>	57
<i>Table 4-5: Regression results of entrepreneurial choices on indices. 29 participants</i>	59

List of figures

<i>Figure 3-1: Timeline of the Ruka Juu and the Social Lab Project</i>	21
<i>Figure 3-2: Background variables from RJ</i>	23
<i>Figure 3-3: Histograms on knowledge</i>	26
<i>Figure 3-4: Time preference</i>	27
<i>Figure 3-5: Competitiveness</i>	27
<i>Figure 3-6: Willingness to take risk</i>	28
<i>Figure 3-7: Self-confidence</i>	28
<i>Figure 3-8: Histograms on Entrepreneurial Ambitions</i>	29
<i>Figure 3-9: Sales and profit per participants from highest to lowest sales</i>	41
<i>Figure 3.10: Total investment per participants from highest to lowest.</i>	41
<i>Figure e 3-11: Average weekly savings from highest to lowest</i>	42
<i>Figure 3-12: Development of weekly averages in the dataset as a whole</i>	43
<i>Figure 3-13: The development of average sales over the year</i>	43
<i>Figure 4-1: Competitiveness and gender differences in sales performance</i>	61

List of abbreviations

BBPT	Berge, Bjorvatn, Pires, & Tungodden
ILO	International Labour Organization
NHH	Norwegian School of Economics
RAWG	Research and Analysis Working Group, Tanzania
REPOA	Research on Poverty Alleviation, Tanzania
RJ	Ruka Juu Project
SL	Social Lab Project
UDEC	University of Dar es Salaam Entrepreneurship Centre

1. Introduction

Perhaps the most fundamental question in experimental economics is whether findings from the lab are likely to provide reliable inferences outside of the laboratory (Levitt & List, 2007, p. 170).

Lab experiments have become common practice in studying behavioural economics, but the studies on the link between laboratory results and how these are extrapolated to the outside world are still in the minority (Benz & Meier, 2011; Levitt & List, 2007). Furthermore, these experimental methods have increased greatly in application (Falk & Fehr, 2003). Various papers study competitive behaviour in a lab setting (Gneezy, Niederle, & Rustichini, 2003; Niederle & Vesterlund, 2007), but the lack of field evidence of how such predictions of behaviour materialise in a real life context was the base of a paper written by Berge, Bjorvatn, Pires, & Tungodden (BBPT)(2014). They aim to contribute to bridging this gap by investigating how competitive behaviour and other entrepreneurial traits emerge in the field behaviour of a set of established entrepreneurs in urban Tanzania. More specifically, they investigate the association that lab results on entrepreneurial traits have to business outcomes, like sales and profits, and to investment and employment choices in the field.

Our paper aims to contribute to understanding the link between the laboratory experiment results and field behaviour. We draw inspiration from and build on the research design, and the empirical strategy of BBPT (2014). The laboratory dataset has been collected in a similar manner, but, although the setting is the same, our field data is sufficiently different. This enables us to extend the validity of the paper in question, and continue bridging the unfortunate gap between laboratory experiments and field behaviour. Our field data is an exceptional set of financial diaries from 49 young entrepreneurs in Dar es Salaam. This data was part of the Social Lab Project by Jacqueline Mgumia, PhD candidate in anthropology. The unique collection method and daily reporting of business outcomes and choices make our field data rich with both quantitative and qualitative information, and more trustworthy because of the frequent reporting.

Investigating young entrepreneurs is highly relevant in development economics. In many developing countries, the lack of official jobs and high unemployment rates leave the youth with few other options than to become self-employed (Bjorvatn, Falch, & Hernæs, 2012a). Policy makers have thus identified entrepreneurship as a possible solution for solving issues

regarding youth unemployment. As a consequence, governments and organizations are aiming to strengthen the entrepreneurs' ability for business survival and growth in developing economies (ILO, 2013). Policymakers have been serving much attention to building up entrepreneurs' human capital, as most self-employed in these economies have low levels of education and business skills. At the same time, development of the right entrepreneurial mind-set and attitudes is increasingly receiving attention in countries like Tanzania.

In this paper, we investigate entrepreneurial characteristics from a laboratory experiment (lab) and how these materialise in the real world (field). We call this the lab-field relationship throughout this paper. The setting of entrepreneurship enables us to study the decision makers in the small-scale businesses. By offering a deeper understanding of the young entrepreneurs in our dataset, uncover their entrepreneurial traits, and describe their challenges, we wish to provide insights into the literature on entrepreneurship in the developing world.

Furthermore, we aim to contribute into the research of Mgumia in two ways. Firstly, by profiling the 49 participants of the Social Lab on different entrepreneurial traits based on lab data. Secondly, by presenting a quantitative and graphical description of the participants' business development over the year based on her field data. This way we may supplement her anthropological work with an economist's perspective.

Quantitatively, we find there is a strong and large association between competitiveness in the lab and business sales and to some extent also profits, and to investment choices. Other entrepreneurial traits, along with knowledge and ambitions to become an entrepreneur, do not seem to have an interesting and logical effect on business outcome variables, or entrepreneurial choices. We also found a strong association between high knowledge and savings.

We point to some of the barriers against the consistency of the lab-field link, by including a qualitative perspective on the topic in question. The contextual constraints that the young entrepreneurs in urban Tanzania face, such as family, cultural and institutional constraints, are barriers for business development. Other constraints are that the young and self-employed often lack education and access to financial capital, like emphasized by the

theories on microfinance. Typically, the focus here is on human and financial capital constraints for business survival and growth.

The authors of this thesis spent 10 weeks in Dar es Salaam, where we also worked as research assistants on the on-going project “Girls Economic Empowerment” by The Choice Lab, NHH. Being there, and working alongside Tanzanian research assistants, and meeting some of the youths from the Social Lab Project, we got an understanding of the culture, and got a feeling of the context of the data that this thesis is built upon.

Outline

The first part of this thesis provides an introduction to existing literature on the lab-field relationship, entrepreneurship and competitiveness, and then finally an introduction to entrepreneurship in Tanzania.

The second part covers the methodology of this study. More specifically, we elaborate on the data sources of respectively the lab and the field dataset and how we use them. We introduce and explain the lab indices, and the key field variables. Finally, we describe the data and the participant’s performance on the key field variables.

The third part sets out to assess the results from a correlation analysis and a regression analysis on the lab indices against the key field indicators.

The paper concludes with reflections on alternative methods, and on limitations and the validity of our results. Subsequently, we provide a summary of the key findings and draw upon existing literature, relating back to the twin study by BBPT (2014).

2. Literature review

The aim of our thesis is to contribute to the understanding of the lab-field relationship. We do this in the context of young entrepreneurs in an urban setting in Tanzania. We specifically focus on the entrepreneurial trait of competitiveness, as this mind-set variable has been proven to have a relevant association with field outcomes and choices.

The literature review is thus constituted of the following literature topics: (i) The Lab-field relationship (ii) entrepreneurship, and, (iii) as this an important entrepreneurial trait, competitiveness, and also including the gender aspect of this topic. Finally, we (iv) present the context, Tanzania.

2.1 Why look at entrepreneurs?

Small-scale businesses mostly operate locally, and not in national economies. Through offering employment, and sustaining many families, they contribute to the bottom of the pyramid. They have been shown to also contribute significantly to national economic growth (Liedholm, 2002). Mensah, Tribe, and Weiss (2007) find that although small-scale businesses in Ghana have a modest income, they are important for sustaining livelihoods and lowering economic risk by diversifying income sources. This implies that focus on entrepreneurship is important for development and economic growth.

Entrepreneurs are also a relevant sample to investigate for the lab-field relationship. BBPT (2014) raise two reasons why. Firstly, business profits can be used as an indicator of success. Secondly, the entrepreneurs are the decision makers in their businesses, and thus highly relevant investigation objects for analysing field choices and behaviour.

2.2 The lab – field relationship

There has been an increase in application of experimental methods on economic questions in the last two decades, and the trend is not likely to stop (Falk & Fehr, 2003). Lab experiments have become common practice in the area of behavioural economics. However, recent articles question the generalization and external validity of such experiments (Benz & Meier,

2011; Levitt & List, 2007). This means that it is important to understand this link, and how lab predictions actually materialise in the field. Is it so that the insights that we gain from the lab can be generalized to also apply in the world outside? Benz and Meier (2011) raise three reasons why people behave differently in the lab. Firstly, when money incentives are involved, the feeling of entitlement to that money is very different in a lab than in the field. Secondly, participants may act the way they do in order to please the experimenter, or because they believe they are supposed to act in a certain way. Finally, they point to the fact that the lab itself is artificial, and lacks the rich real life context. They, however, find that there *is* a correlation between donation behaviour in the lab and charitable giving in the field two years before and two years after the actual experiment. Other researchers also find evidence confirming this link, while others question it.

Levitt and List (2007) call for recognizing the shortcomings in experimental research. The laboratory experiments in physical science are similar to the ones in experimental economics. However, for the latter, there are so many additional varying factors affecting human behaviour in the lab compared to the outside world. They raise particular focus to five of these factors, which are moral considerations, the type and degree of other's scrutiny of one's action, the context, the self-selection of the individuals making the decision, and the stakes that the individual is up against. Despite the constraints, experiments can be designed to help combat them by anticipating the biases, and by combining lab experiments either with another lab experiment, with a field experiment or with theory. Similarly (Falk & Fehr, 2003) discuss the pros and cons of experiments in the context of labour economics. They also propose ways of bypassing the problems, and that lab and field experiments are very useful instruments as complements to traditional methods. Although both these papers point to the deficiencies of experimental methods in economics, they still believe in the usefulness given that the experimenters take certain measures in combating the flaws. This shows the importance of understanding the link between the lab and the real world in order to be able to design experiments carefully so that they may be extrapolated to the outside world.

Both Zhang (2013) and Buser, Niederle, and Oosterbeek (2014) investigate the relation between standard experimental measures of competitiveness and a student's education decisions. They find evidence that measures of competitiveness is relevant in predicting choices and performance in the field, such as the choice of future academic tracks (Buser et al., 2014), and relevant for the probability of taking competitive entry exams (Zhang, 2013).

Sutter, Kocher, Gltzle-Rtzler, & Trautmann (2013) find that experimental measures of impatience predict worse conduct at school, a lower probability of saving and undesirable health-related field behaviour (such as a higher body mass index, and higher spending on alcohol and cigarettes). Their experimental measures for risk and ambiguous attitudes only slightly predict behaviour.

BBPT (2014), as mentioned, investigate how competitive behaviour in the lab is associated with field choices and outcomes. In the setting of entrepreneurship in Tanzania they specifically look at the correlation between competitiveness in a lab experiment and competitive choices and successful outcomes in the field. They find competitiveness to be an important entrepreneurial trait to affect field choices, and, to some degree outcomes.

2.3 Entrepreneurship

The literature on entrepreneurship has many views on which characteristics seem to be of more importance for business survival and growth. From the classical literature, we find Joseph Schumpeter's definition from 1934 of an entrepreneur as an innovator that introduces new or improved products or solutions into markets has been widely recognised (Backhaus & Schumpeter, 2003). Another view is represented by Lazear (2005), which argues that entrepreneurs need the ability to combine human, physical and informational resources simultaneously and efficiently. He argues that entrepreneurs with sufficient skills in a *variety* of areas, what he calls "Jacks-of-all trade", have a competitive advantage when creating a successful business.

In the setting of small-scale entrepreneurs, recent literature on microfinance points to the lack of human capital among the entrepreneurs, in the form of business knowledge, managerial skills and financial literacy, as a restricting factor for growth and success in the business. BBPT (2014) find that a financial grant to small-scale entrepreneurs in Tanzania only has an effect on business outcomes when offered in tandem with business training. The effect of financial capital has been shown to depend on educational background, business skills and mind-set (Duflo, Banerjee, Glennerster, & Kinnan, 2013). Human capital has been shown to be a determining factor for growth and survival among small enterprises in Africa and South America (Liedholm, 2002).

Neneh (2012) looks at the importance of having an entrepreneurial mind-set for business success in a South African context, and claims that the lack of such a mind-set is one of the reasons for the high failure rate of small and medium-sized enterprises (SMEs), and that most great and successful leaders have a growth mind-set. Dhliwayo & Van Vuuren (2007) write that there is no difference between an entrepreneurial mind-set and strategic thinking, and by this, that entrepreneurial mind-set is important in order to succeed. Kroon (2002) (in Neneh, 2012) revealed a strong relationship between a willingness to take risk and entrepreneurial business success.

Also de Mel, McKenzie, & Woodruff (2008) look at entrepreneurial characteristics. They argue that for a substantial part of the microenterprises, the slow growth is likely to derive from a lack of ability or desire to grow rather than a shortage of finance. Their analysis concludes that only a third of microenterprise owners have characteristics typical for entrepreneurs. The factors that most clearly differentiate them are cognitive ability, motivation and competitive attitude. Arriving from this, we see that skills, ambitions and a competitive attitude seem to be important entrepreneurial traits to stimulate growth. Enhancing these traits, and understanding their relation to successful outcomes and entrepreneurial choices, is thus likely to be important.

2.3.1 Competitiveness and gender differences

As mentioned, BBPT (2014) find a strong correlation between competitiveness and business outcomes and choices. The growing literature on competitiveness is specifically concentrated around gender differences observed in competitive settings.

The experimental literature has identified significant gender differences in willingness to compete, and also on performance in a competitive environment (Gneezy et al., 2003; Niederle and Vesterlund, 2007). Evidence from experimental measures of competitiveness could explain gender differences in educational and occupational choices in the real world (Almås, Cappelen, Salvanes, Sørensen, & Tungodden, 2014).

In their study, Gneezy et al. (2003) conducted a controlled experiment to analyse how performance in a competitive environment might differ between genders. In an experimental design study, participants were set to solve computerized mazes in both a non-competitive and competitive environment, and in both single- and mixed-sex groups. The participants were further paid for their performance either for each maze solved, or in a winner-takes-it-

all tournament. The main result shows a significant increase in the performance of the male participants when the competitiveness in the environment increases, but no effect among the female participants. In the non-competitive environment, little or no gender differences were reported. However, in the single-sex tournament, female performance was significantly higher.

In a similar study, Niederle and Vesterlund (2007) examine whether male and female differ in their willingness to *enter* a mixed-sex competition. They find that the majority of females choose the non-competitive alternative. Moreover, controlling for individual ability, their evidence shows that many high-performing females thus limit themselves financially by choosing not to compete. The opposite is proven for the male participants, where too many poorly performing males choose to enter competition. Furthermore, this indicates that women might not perform to their maximal abilities in the mixed-sex competitive setting.

Experimental literature, as we see, suggests that men tend to outperform women in an experimental competitive environment, and more frequently select the competitive alternative than women. However, as written by Croson and Gneezy (2009), women that choose a competitive environment seem to perform at a similar level as men.

In the case of the lab-field relation of, the findings are much more ambiguous when it comes to gender differences in competitive settings. On one hand, Buser et al. (2014) find significant evidence of gender differences in the choice of academic tracks in the Netherlands. On the other hand, Zhang (2013) finds no such difference in her study of competitive inclination in China. Literature that suggests women have a lower preference to competitive environments is still well supported, however (Croson & Gneezy, 2009). The explanations for why these differences occur are many, and both biological factors and environment may explain the gender differences in competitive settings (Bjorvatn, Falch, & Hernæs, 2012b)

2.4 The Context. Youth and Entrepreneurship in Tanzania

Tanzania- a brief introduction

The United Republic of Tanzania was constituted in 1964, after the two former British colonies Tanganyika and Zanzibar merged after achieving their independence in the early

1960s. The country is located in East Africa and has a population of about 49 million. Tanzania has a young population with about 45% being under the age of 15 (Central Intelligence Agency, 2014). It is also a fast growing population, at a yearly growth rate of 2.8%, due to high fertility rates and increased life expectancy (Central Intelligence Agency, 2014).

Over the last decade Tanzania has experienced a steady economic growth with an annual GDP growth rate of 6-7%, which is predicted to stay anchored at high-level rates in the years to come, particularly in the urban areas (World Bank, 2014). This growth can be accredited to the fast growing capital-intensive sectors. However, the economy is still heavily dependent on agriculture sectors, which employs 80% of the Tanzanian workforce and accounts for 85% of the country's exports. This leaves Tanzania highly vulnerable to volatility in commodity prices and international markets (Central Intelligence Agency, 2014).

Like many developing countries, Tanzania is experiencing a rapid urbanization at an annual rate of 4.77% - almost twice the size of the total population growth (Central Intelligence Agency, 2014). This rural-to-urban migration is mainly dominated by young people aged 15-35 (REPOA, 2014). The previous capital, Dar es Salaam, is the largest and fastest growing city. After the capital was moved to Dodoma in 1973, Dar es Salaam still remains the most important economic and administrative centre in the country. In a recent report, The World Bank projects more Tanzanians to live in urban areas and the population of Dar es Salaam to exceed 10 million by 2030. The current urbanization process has positive impacts on the transformation of the economy and the labour force. The contribution of non-agricultural sectors to the GDP is increasing (World Bank, 2014 -b).

Despite overall high economic growth rates, Tanzania still remains a poor country, and ranks 154 on the Human Development Index ranking. Through the Tanzania Development Vision 2025, the country strives to reduce poverty through economic growth, and achieve middle-income country status (UNDAP Tanzania, 2013). However, Tanzania is still far from meeting the poverty reduction goals, and there are high concerns about the ability to create

enough productive jobs for the rapidly growing workforce¹. According to the World Bank (2014 -b) the Tanzanian workforce increases with about 800.000 youths every year.

2.4.1 Education in Tanzania

High investments in primary education have been an essential part of the Tanzanian government's development strategy and step toward reaching the Millennium Development Goal number 2². Introducing free education in 2001 and making it mandatory to send your children to primary school are the main reasons why 95% of school-aged children in Tanzania had access to primary education in 2009 (United Nations in Tanzania, 2014).

Primary education consists of seven years, while the secondary level is divided into ordinary level from Form I-IV and advanced level from Form V-VI, with a total of six years. As opposed to primary school, secondary education is optional and enrolment rates are a lot lower. Two national exams are demanded as qualifications to advance to higher levels of education. The Primary School Leaving Examination (PSLE) is taken at the end of primary school to advance to ordinary level. The Certificate of Secondary Education Examinations (CSEE) at the end of Form IV to qualify for advanced level. The exam fees are generally expensive and many students are therefore not able to get hold of their results, so that they may proceed to the next education level (2014). Despite high enrolment rates in primary level, the quality of the education is questionable, mostly due to the lack of qualified teachers and the rapid enrolment growth (United Nations in Tanzania, 2014). As a consequence, more than half of the students fail their exams³. The high drop out rate, underlines the pressing quality issues in the Tanzanian educational system.

1 Today approximately 23 mill are in the job market, and by 2030 the estimate is 45 million (World Bank, 2014 -a).

2 Millennium Development Goal number 2 is to achieve universal primary education.

3 According to Ministry of Education and Vocational Training (2012), barely 50% of the students passed the PSLE in 2009, down from 70% in 2006.

Entrepreneurship in urban areas

The lack of job opportunities leaves the majority of youths entering the labour market with no other option than to become self-employed and start small enterprises⁴. Most of these are informal, which means that they are not registered and do not pay taxes, and normally have few or no employees. The informal enterprises are a significant source of employment for the urban youth in Tanzania (REPOA, 2014). These forms of enterprises often function as a stepping-stone into employment or establishment into the formal sector⁵. A large proportion of the youth lack proper education and training, and they often have low productivity capabilities and low-income levels. This limits their capacity to expand their enterprise and become competitive⁶.

In addition, the Tanzanian Research and Analysis Working Group (RAWG) states that high levels of rural-to-urban youth migration will result in an increased level of informal urban enterprises, where the labour productivity is low. Further, RAWG states that these informal enterprises will limit the prospects of the country to develop a more diversified economy, as they are excluded from the formal mainstream economy (REPOA, 2014).

The transition from informal to formal sector is particularly challenging for the youth to achieve, as it involves requirements such as having a business licence and a business premise, and the process is often too costly and complex for these entrepreneurs (ILO, 2013). The main constraints that urban small enterprise owners face are lack of required skills, lack of access to external finances, high costs of connectivity, insecure administrative environment and weak rule of law (World Bank, 2014-c).

Emphasising the improvement of skills, and access to finance, technology information, business practices and markets can for small enterprises have a positive impact on growth (World Bank, 2014-c). During the recent years the Tanzanian government, organisations and

⁴ Urban unemployment rates are both consistently and significant higher than the national average of 17.5% (CIA, 2014). The unemployment rate in Dar es Salaam for those aged 15 and above was 31.5%. For the youth it is significantly higher; those aged 15-24 at 14.5% and 25-34 at 11.8%. (2014)

⁵ The formal sector includes all jobs with normal hours and regular wages, and is recognized as income sources on which income taxes must be paid (Business Dictionary, 2014)

⁶ As much as 93.2% of the self-employed workers without employees operate in informal sector (2014).

universities have recognized entrepreneurial training programs as a tool to increase labour productivity and to facilitate growth in these small enterprises.

3. Methodology

We base our thesis on a rich set of data, ranging from lab experiments, financial diaries, and field interviews.

In Section 3.1, we first describe the lab data from the Ruka Juu project (RJ). We describe the participants on their general background and socioeconomic status, and compare them to the rest of the RJ sample. Thereafter, we introduce the lab indices and describe how our pool of participants scores on these indices. We then again compare them to the RJ sample for generalizability. In Section 3.2, we describe the field data based on financial diaries collected as part of the Social Lab throughout 2013, and field visits in 2014.



Figure 3-1: Timeline of the Ruka Juu and the Social Lab Project

While BBPT (2014) investigate 207 already established microfinance entrepreneurs at an average age of 39, our paper investigates the characteristics of a set of 49 young Tanzanians at start-up, their average age at the time of the RJ baseline being 18. All 49 were previously part of the Ruka Juu experiment in 2011 and were selected on two criteria, (1) that they were not in school at the beginning of the Social Lab Project (SL) year of 2013, and (2) that they showed interest when they were asked in the RJ survey whether they would like to use their 4000 TZS⁷ participation reward to pay for two weekend-long courses on finance and/or

⁷ To give the reader an idea about the monetary measurements provided in this thesis, we give a little description of the exchange rate of the Tanzanian shilling, the cost of living in Dar es Salaam and the purchasing power. One thousand Tanzanian shillings (TZS) was the equivalent of US\$ 0,63 on January 1 2013, or the inverse of 1 581 TZS for one dollar. One thousand TSZ was about 3,5 NOK (Norwegian Kroner), or the inverse of 284 TZS on one NOK (XE, 2014). A litre of milk costs about 2 500 TZS, a loaf of fresh bread would be 1 500, rice at 2 000 and a kilo of tomatoes at 2 300. The cheapest type of meat or fish would maybe be around 8 000 on the market. The national poverty line per capita per month in Tanzania is US \$19 (in 2005 dollars, and adjusted for purchasing power) (Ravallion, Chen, & Sangraula, 2009). The national poverty line is calculated based on the cost-of-basic-needs method. It is a monetary estimation of the cost of a food and non-food bundle that is regarded absolutely necessary to sustain human life (Bundervoet, 2013). Private consumption expenditure according to the National Accounts was \$45 (Ravallion et al., 2009).

starting a business⁸. In addition to the criteria of interest in training, none of the 49 participants advanced to A-level (Form V) after completing Form IV.

3.1 The Lab Data

In the following, we (1) explain the source of the lab data, the RJ project; (2) describe the general and socioeconomic background of our participants; and (3) describe the collection of and reasoning behind our key entrepreneurship indicators. We profile our pool of participants based on these indicators, and then (4) compare them to the rest of the RJ pool on both the background variables and the lab indices. If they are similar to the RJ pool, the results from this paper may be more generalizable to other Tanzanian youth.

3.1.1 The Ruka Juu Project

The RJ experiment from 2011 was an evaluation of an educational entertainment (edutainment) TV show in Tanzania. The project was a randomized control trial (RCT) where they looked at whether the show could inspire young people to start their own business and teach them how to do so (Bjorvatn, Cappelen, Sekei, Sørensen, & Tungodden, 2013).

The dataset describes the socioeconomic background of 2136 participants and their level of knowledge, entrepreneurial mind-set, and ambitions to become entrepreneurs. Participants answered both incentivized and non-incentivized questions, and were not given any feedback on their performance during the experiment. The follow up survey data from June 2011 forms the lab data of our thesis, and the socioeconomic background information is taken from the baseline.

In order to be able to reach a large number of participants, the pool of participants for the RJ experiment were all in school. They randomly chose 43 schools in the Dar es Salaam region. The subject pool was all at the secondary school level in the last year of Ordinary level (Form IV). In the following we restrict ourselves to describing the dimensions used in this

⁸ 12% of the treatment group and 8% of the control group showed interest in further business training after the RJ. This training was later given by University of Dar es Salaam Entrepreneurship Centre (UDEC) to a randomized selection (Mgumia, 2014).

paper, but for details about the experiment and the treatment effects, see (Bjorvatn et al., 2013).

3.1.2 The general and socioeconomic background

We use the participants' gender, educational stream, socioeconomic background and whether the head of house runs his own business as control variables in our regressions in this paper⁹. Falk & Fehr (2003) point to the importance of including controls. A growing number of studies find that gender has an effect on business outcomes (Buser et al., 2014). We include educational stream as a background variable, as type of stream indicates their expertise. Further household socioeconomic status is included, since parental socioeconomic status can effect the children's future education and labour market outcomes (Leppel, Williams, & Waldauer, 2001). Finally, if the house head runs his own business, this can be of inspiration and help to the child. We chose these background variables as controls, as they might influence the outcomes and their choices.

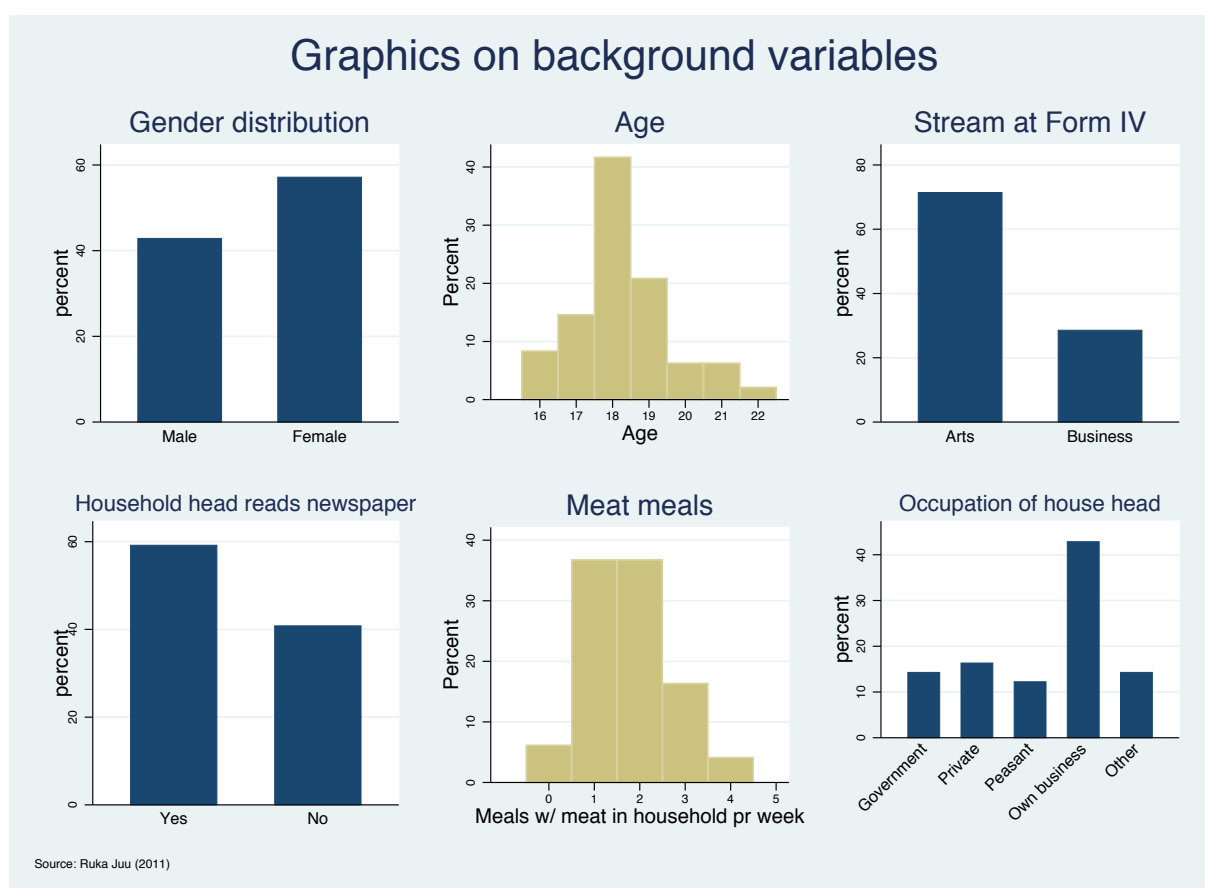


Figure 3-2: Background variables from RJ.

⁹ We do not include age, as they are of the same age group.

The participant pool of 49 that we follow in this paper consists of 21 boys and 28 girls, and even though all were in Form IV during the Ruka Juu Project, the age difference between the youngest and oldest participant is six years. The average age of the 49 participants is about 21 years (in 2014). About 71% took arts as education stream during Form IV, while the rest took business stream.

The socioeconomic variables are measures of household sophistication level (if household head reads newspaper) and household wealth (by number of meals with meat per week). 59% of the participants have head of household reading newspapers, and they eat meat on average 1.7 days a week. In 40% of the households, the head of house runs his own business. The rest are employed, or do other things.

Table 3-1: Background summary statistics from RJ 2011

	Mean	Std. Dev.	Min	Max	Obs
Gender (M/F)	1.57	.5	1	2	49
Stream (Business or Art)	1.29	.46	1	2	49
Newspaper (Y/N)	1.41	.50	1	2	49
Meat in a week	1.76	.947	0	4	49
House head occupation	3.27	1.30	1	5	49
Age	18.30	1.34	16	21	48

3.1.3 The entrepreneurship indicators

We base our choice of entrepreneurial indicators on established studies within the field of entrepreneurship, and are particularly inspired by the Ruka Juu study itself. In line with (Bjorvatn et al., 2013), we use knowledge, entrepreneurial mind-set and ambitions to become an entrepreneur as the determinants of entrepreneurship. BBPT (2014) also use a similar set of variables to capture the dimensions of an entrepreneur. We make three main indices, each based on four questions (sub-indices) in the RJ follow-up survey. The main indices are knowledge (*HIKnow*), entrepreneurial mind-set (*HIMinds*) and ambitions to become an entrepreneur (*HIAmb*). Variable names and indices from the lab are given in italic throughout this paper. The purpose of the main indices is to capture a broader set of dimensions in each of these, and to see if the aggregates have a different impact on outcomes and choices. For knowledge, we wanted to be able to see the effect of the combination of mathematical skills and business knowledge, in line with Lazear's concept of an entrepreneur as possessing a broad set of knowledge and skills. For mind-set, we wanted to

also see the combined effect, and if one characteristic could compensate another, for instance if a lack of competitiveness could be compensated by a high willingness to take risk. For ambitions, we wanted an indicator that has a broader range, and that captures more than just interest in training or in owning your own business.

In the following we describe how the data was collected, and the aggregation method used for each of the main indices. We, however, put more emphasis on the mind-set variables because these are of particular interest, since an entrepreneurial mind-set is important for growth. We want to investigate how these dimensions affect outcomes and choices both aggregated and separately. We also want to understand these dimensions so we can use them for comparison to previous lab-field studies, like BBPT (2014) and Sutter et al. (2013).

Knowledge

The measurement on knowledge is based on four variables from the RJ dataset, namely business knowledge, math grades, calculations and entrepreneurship knowledge. See the scores of our pool of participants in Figure 3-3 and in Table 3-2.

First, to measure business knowledge (*know*), the participants were asked 16 questions, like what profit is, when insurance is most useful, and how much repayment is required to the bank given certain information. Math grades (*mathp*) were collected from the schools where the surveys were conducted. To measure calculation skills (*cadd*), the participants added numbers 83 times. Finally, entrepreneurship knowledge (*entp*) was measured given questions like what is an important element in a business plan, how do you calculate your sales for a particular good, which businesses need to be registered in Tanzania, etc.

Within each sub-category of the knowledge index, we divide their personal result by the mean result of the whole Ruka Juu data set. This is to get a sense of the relative score compared to the approximately 2000 other participants. For instance, for a score of 1.2 on business knowledge, the participant did 20% better than the average for the whole RJ sample. A score of 1 means average etc. Finally, we make an average out of these four scores for each participant, which makes their knowledge index score (*HIKnow*).

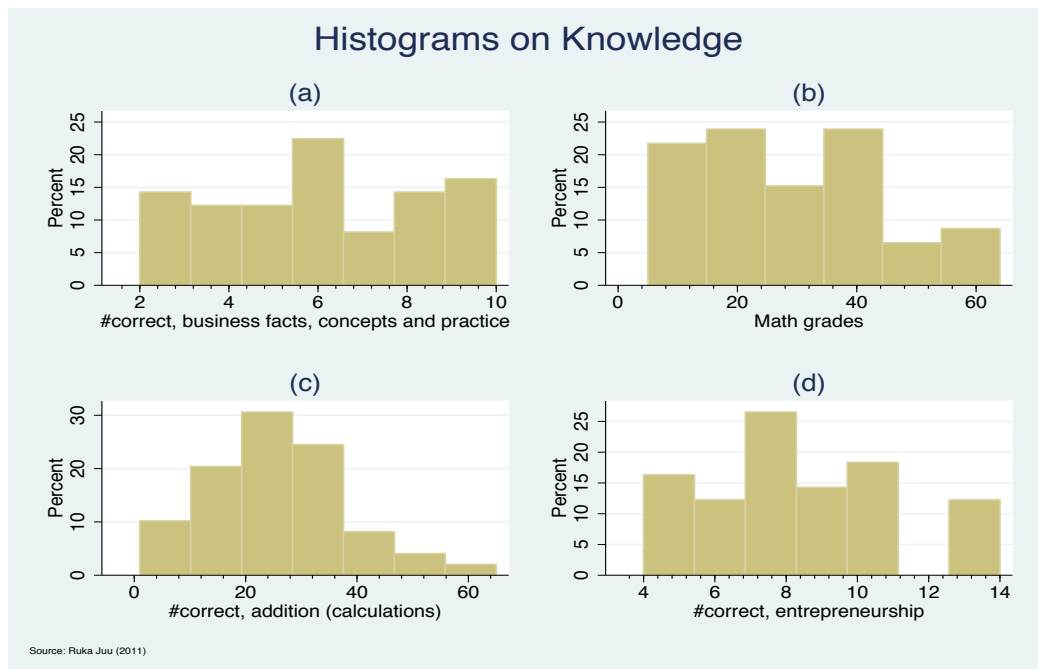


Figure 3-3: Histograms on knowledge. Graphs on our participant pool of 49. Graph (a), the knowledge graph, shows the result distribution on the business knowledge. Graph (b) is the distribution of the math grades in per cent. Graph (c) shows the distribution of the results on the calculation test. Graph (d) shows the distribution of the results from the entrepreneurship test.

For our pool of participants, the average of *HIKknow* is 0.95, which means that they score lower than the average RJ participant. The lowest score in our pool is at 0.46, which is 54% below the average score. The participant with the highest score does 50% better than the average RJ participant.

Table 3-2: Knowledge summary statistics

	Variable	Mean	Std. Dev.	Min	Max	Obs
Knowledge & Skills	<i>HIKknow</i>	.95	.26	.46	1.50	49
Business knowledge	<i>know</i>	6.10	2.16	2	10	49
Math grade	<i>mathp</i>	29.57	15.10	5	64	46
Calculation skills	<i>cadd</i>	25.59	12.74	1	65	49
Entrepreneurship knowledge	<i>entp</i>	8.33	2.78	4	14	49

Entrepreneurial mind-set

To create the index for mind-set, we use the same four non-cognitive factors from lab as BBPT (2014), which are (1) time preference, (2) competitiveness, (3) willingness to take risk, and (4) self confidence, and call them the entrepreneurial mind-set sub-indices. See the summary statistics of the scores of our pool of participants in Table 3-3.

Firstly, time preference, or patience, (*patient*) was in the lab measured by three cases in two rounds. First they were given three cases where they could choose between receiving 1000 TZS today, or in 8 weeks receive 1500, 3000 or in the last case 5000 TZS. In the next round they were given the same cases, but the time

frames were changed to 8 and 16 weeks, respectively. The reward in 16 weeks was still incremental. The most patient participants chose the alternative in the future all six times. Our pool of participants has, on average, chosen to receive more money in the future 3.69 times, so a little over half of the time. Most of them (36.73%) chose to wait for a higher award four out of six times. See figure on Patience.

Secondly, competitiveness (*competitive*) was measured by asking if participants wanted to compete in the second round of calculations (see discussion about knowledge). They could choose whether to work at a fixed rate of 100 TZS for each correct answer, or at a competitive rate where they received 300 TZS for each correct answer only if they

performed at least as good as the average in the first calculation round. Only 13 out of 49 participants chose to compete in the RJ lab competition. The figure on the Compete-sample shows the share of how many chose to compete and how many did not.

Thirdly, willingness to take risk (*risk*) was measured in three cases, where the participants chose between a safe (2000 TZS) and a risky alternative (4000 TZS). The choice was the same in each case. However, the probabilities of the lucky outcome in the risky alternative were 25%, 50% and 75%. The most risk-willing would then choose the risky alternative in all 3 cases. In our pool of participants, on average they chose the risky alternative 1.4

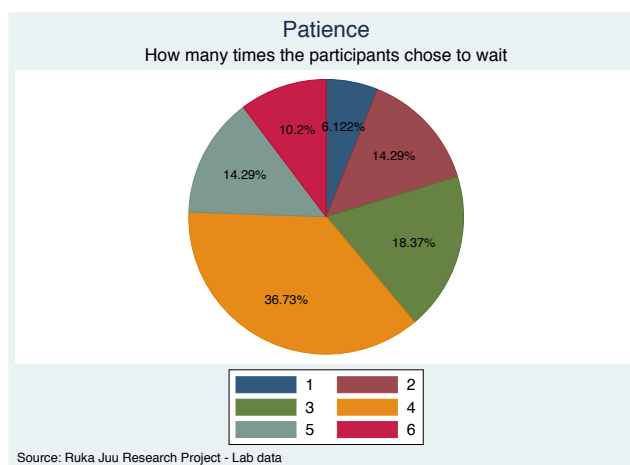


Figure 3-4: Time preference

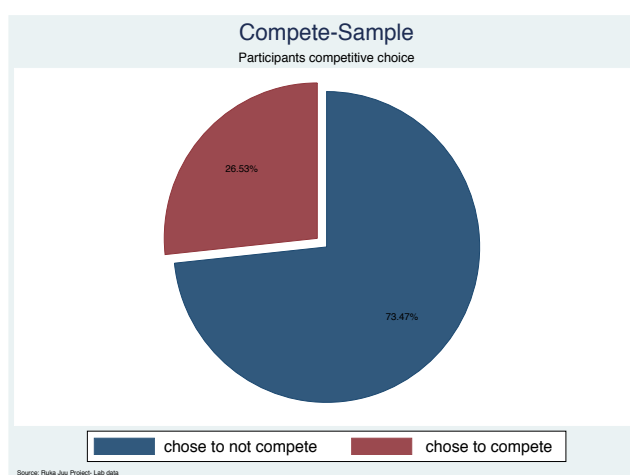


Figure 3-5: Competitiveness

times, so they are medium risk-willing. The majority chose the risky alternative either 1 (44.9%) or 2 (38.8%) out of 3 times. See figure on Willingness to Take Risk.

Fourthly, self-confidence (*confidence*) was measured in advance by asking the

participants how they expected to perform in the first calculation test in comparison to the other participants in the experiment. The most confident answered that only 0-10% would do better than them, and the least confident chose alternative 1, that 90-100% would do better. 24% of the participants in our dataset believed that only 0-10 % would do better than them. Out of 10 options, where 10 indicates the most confident, the participant average is 6.7, which suggests that they generally consider themselves better than average in calculations. See figure on Confidence.

Finally, to create aggregated, main mind-set index (*HIMinds*) the participant scores one point for every one of the four sub-indices that are well above the average of the whole

RJ dataset. These points are aggregated up into the mind-set index measured from 0-4 where 4 is the highest score. Scoring 4 on the *HIMinds* index thus indicates that the participant is more competitive, more willing to take risk, more patient and at the same time is more self-confident than the average RJ participant. The average for our pool of participants is 1.7.

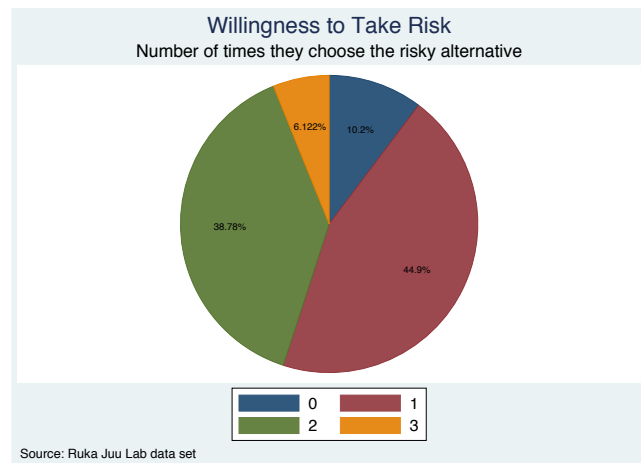


Figure 3-6: Willingness to take risk

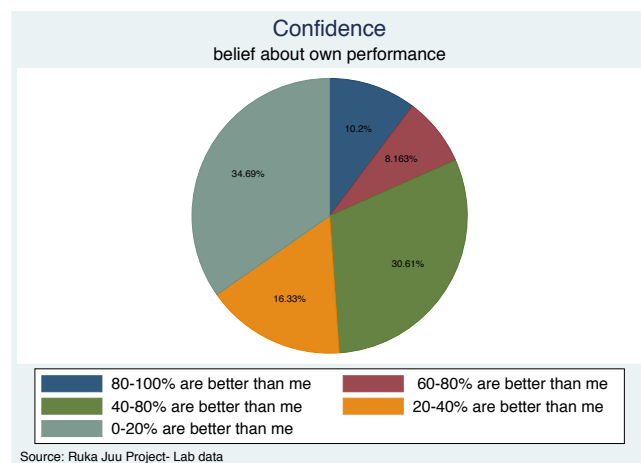


Figure 3-7. Self-confidence. In the RJ the intervals between each alternative were on 10% intervals (i.e. 0-10% are better than me). In this graph we have aggregated to 20% interval level for a neater graphical presentation. We see that over 50% of our pool of participants believes that 0-40% are better than them

Table 3-3: Mind-set summary statistics

	Variable	Mean	Std. Dev.	Min	Max	Obs
Mind-set	<i>HIMinds</i>	1.75	1.01	0	4	49
Time preference	<i>patient</i>	3.69	1.34	1	6	49
Competitiveness	<i>competitive</i>	.26	.45	0	1	49
Willingness to take risk	<i>risky</i>	1.40	.76	0	3	49
Self-confidence	<i>confident</i>	6.78	2.67	1	10	49

Ambitions to become an entrepreneur

“Ambitions to become an entrepreneur” is measured by the most entrepreneurial answer(s) to four survey questions. They were given a score of 1 for every time they chose the most entrepreneurial answer. The responses were aggregated up into an entrepreneurial ambitions index (*HLAmb*) measured from 0-4, where 4 is the highest score.

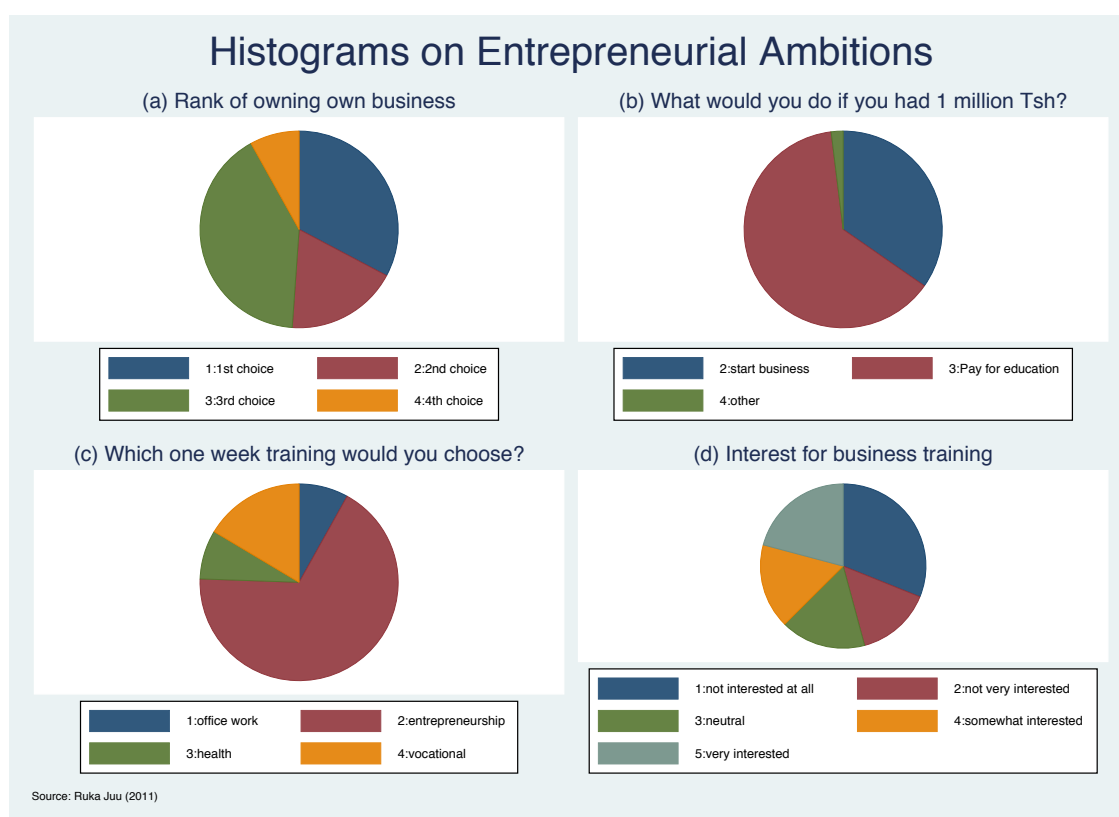


Figure 3-8: Histograms on Entrepreneurial Ambitions. Graph (a) shows that 51% of the participants ranked owning a business over other options as 1st or 2nd choice. In graph (b) we see that 31 % said they would start a business if they got 1 million TZS. Graph (c) illustrates that 67% would prefer the topic of a one-week training to be on entrepreneurship, and in graph (d) we see that 37 % were very or somewhat interested in business mentorship

First, they gave ranks on having their own business (*rank*) over being a private or government employee or a farmer. The most entrepreneurial answer was to rank this as 1 or 2. Second, they chose between four different options of what they would do if they had 1 mill TZS, where the most “entrepreneurial answer” was to invest in starting a business (*startb*). Other options were, for example, further studies, or to buy something nice for the family. Third, they chose between four different topics in a hypothetical free, weeklong training course, where the most entrepreneurial answer was “training in entrepreneurship” (*tr_ent*). The other options were office work, health and vocational training. Finally, they reveal how interested they are in mentorship with a business-person (*mentor*) on a scale of 1-5, where the most entrepreneurial answers were 4, “somewhat interested”, or 5, “very interested”.

The *HIAmb* index value is thus the sum of the indicator variables with an optimal value of 4. The average of our pool of participants is 2.3. See Table 3-4 for the summary statistics.

Table 3-4: Ambitions summary statistics

	Variable	Mean	Std. Dev.	Min	Max	Obs
Ambitions to become entrepreneur	HIAmb	2.30	1.02	1	4	49
Rank of having own business	rank	2.24	1.01	1	4	49
Preference of starting own business if has 1 million TZS	startb	.35	.48	0	1	49
Would choose entrepreneurship training	tr_ent	.67	.47	0	1	49
Interest for business training mentor	mentor	2.81	1.55	1	5	48

3.1.4 Comparing our pool of participants to the RJ dataset

First, we compare the background of our participant pool of 49 youths with (1) the whole RJ dataset and (2) those from the RJ that chose to invest in further training programs. This is done to get an idea whether the background of the pool of participants is any different from the rest of the general survey sample. Similarity in the different pools suggests that results from our analysis should be applicable for the other pools as well. We use graphical and summary statistics for comparison. Secondly, we do a similar comparison of the

participant's performance on the lab-indices, *HIKnow*, *HIMinds*, *HIAmb*, and the sub-indices. As with the background variables, graphical, descriptive and summary statistics are used for comparison. In addition, we supply with formal statistical tests on the lab indices.

Background variables

We find that a higher percentage of our participant pool studied arts in Form IV (see section 3.1.2). The head of their households are more likely to be reading the newspaper on a general basis, and they eat fewer meals that contain meat in a week. This indicates the head of the households of our participants are slightly more educated and sophisticated, but that they come from less wealthy households compared to the whole RJ dataset. Other than this, their background seem to be quite similar to the rest of the youth from Dar es Salaam in their age group, as the participants of the whole RJ may be interpreted to be average since they were randomly chosen. We get similar results comparing with the rest of those who chose training in the RJ lab.

Table 3-5: Background summary statistics from RJ 2011

Background variables	49 participants		Whole RJ		Willing to invest, RJ	
	(A) Mean	Std. Dev.	(B) Mean	Difference A-B	(C)Mean	Difference A-C
Gender	1.57	0.50	1.55	0.02	1.58	0.00
Age	18.29	1.34	17.98	0.31	17.91	0.38
Stream(Bus. or Art)	1.29	0.46	1.39	-0.10	1.39	-0.11
Newspaper	1.41	0.50	1.60	0.15	1.28	0.13
Meat	1.76	0.95	2.14	-0.39	2.13	-0.38
Occupation	3.27	1.30	3.17	0.09	3.22	0.05

Lab indices

A way of comparing groups is to run t-tests. These require the data to be normally distributed. We therefore ran a skewness and kurtosis test, along with a Shapiro-Wilk normality test (see Appendix A) on the lab indices for our pool of participants, only the knowledge indices could not be rejected as being normally distributed, and we therefore ran a regular t-test on those. The test conveys if there is any statistical difference between our pool of participants and the rest of the RJ participant pool.

For the mind-set indices that are not normally distributed, we complement with a Wilcoxon rank sum test, which is a nonparametric alternative to the t-test on ordinal data (Lærd Statistics, 2014). *Competitive* and the ambition indices can, however, not be ranked, as they

are dichotomous data (yes/no) and we therefore use a proportion test to check for a significant difference in proportions between two groups (Gerald & Keller, 2012, p. 487). This would then show whether the proportion of competitive people in our participant pool is significantly different from the proportion in the rest of the RJ population.

Compared to the whole RJ dataset, we find that our pool of participants score lower on *HIKknow* at a 10% significance level, which is mostly caused by low calculation (*cadd*) performance. The latter is significant at a 5% level. Based on the summary statistics, we find a difference in willingness to compete, where 37% of the whole RJ dataset chose to compete, while only 26% of our pool of participants did the same. The proportion test shows that this difference is significant at a 10% level. For the rest of the mind-set indices, there are no significant differences. Of the ambition indices, only the interest in training seems to differ. Our pool of participants seems to be more ambitious than the whole RJ sample at a 10% significance level.

To sum up, we find that our pool of participants has a slightly lower knowledge level, and is less competitive, but they are not significantly different in terms of ambitions from the whole RJ dataset.

Comparing our participant pool to the RJ group that chose to invest in training, we find that they are similar in *HIKknow*, and that they are not statistically different in terms of *competitive*, nor the other mind-set indices. Our pool of participants is more ambitious. A higher proportion wanted business mentoring (significant at a 1% level), and they are significantly more interested, at a 10% level, in entrepreneurship training and in starting their own business. To sum up, our pool of participants is similar to those who chose training on *HIKknow* and the mind-set indices, but are more ambitious.

Table 3-6: Lab indices summary statistics from RJ 2011: A presentation of the difference in the mean values between our pool of participants and the other two groups.

	49 participants		Whole RJ		Willing to invest, RJ	
Lab indices	(A) Mean	Std. Dev.	(B) Mean	A-B	(C) Mean	A-C
HIKknow	0.95	0,26	1,00	-0,05	0,98	-0,03
patient	3.69	1,34	3,42	0,27	3,62	0,07
competitive	0.27	0,45	0,37	-0,10	0,30	-0,04
risky	1.41	0,76	1,36	0,05	1,41	0,00
confident	6.78	2.68	7,10	-0,32	7,04	-0,27
HIAmb	2.31	1.02	2,06	0,24	1,89	0,41

Alongside the discussions of the comparison tests above, we see in Table 3-6 that the differences in the means between the groups, particularly the others that were willing to invest in further training, are overall relatively low. This may therefore indicate that our results to some extent also can be applicable to other samples.

3.2 The Field Data

In the following, we will (1) explain the source of the field data, the Social Lab project, (2) describe the cleaning process of the field data to reassure quality, (3) describe the collection of and reasoning behind our key field variables, and (4) portray the participants' businesses and performance over the project year, 2013.

3.2.1 The Social Lab Project

The Social Lab was a project constructed as part of the PhD study "*Choice of Money: Entrepreneurship, Livelihood and Youth Aspiration in Tanzania*" by Jacqueline Mgumia from December 2012 to December 2013. The aim was to look at the relationship between the nature of entrepreneurship training and business start-up processes among youth in the urban setting of Dar es Salaam¹⁰. Mgumia followed them closely throughout the start-up year of their business, had frequent meetings with them, and even stayed with some of them in their homes. The project uses a research technique called financial diaries. The financial diaries were first used to study money management methods and financial behaviour among poor households (Collins et.al 2009). In our case, we capture the financial flows of 49 youths given a business grant of 200 000 TZS, and with limited access to formal education and employment.

The participants were divided into four groups according to their treatment status in previous studies. Group A's 13 participants received both Ruka Juu and UDEC treatment, Group B's 11 participants received UDEC treatment, group C's 13 participants received Ruka Juu treatment, and group D's 12 participants received no treatment. Over the year, they all took part in monthly meetings with their groups, and kept individual financial diaries of their personal economy and business income statements, among other things (see Appendix B).

¹⁰ Personal communication with Mgumia

The financial diary is reported with weekly intervals, and provides insights into participants' financial behaviour and performance on both a daily and weekly basis. Their behaviour is reported on two levels: firstly on a personal level through information on personal incomes, and personal and community/family expenditures; secondly, on a business level, through sales and running costs, investment costs, stock, loss and profits. In addition, the financial diary provides information about time spent on daily activities, a happiness index and quantitative information about weekly plans, thoughts, experiences and challenges.

Table 3-7: Key field variables summary statistics

Key field variables <i>Summary statistics</i>	Mean	Std. Dev.	Min	Max	Obs
Business survival	.59	.50	0	1	49
Average weekly business sales*	90 049	205 655	0	1 306 739	49
Average weekly profit	5831	39 719	-178 731	146 510	49
Total investments over the year	78 914	136 206	0	640 000	49
Average weekly business purchases	76 554	188 448	0	1 206 116	49
Average weekly savings	58 368	93 737	0	558 919	49

*Excluding participant 2 on sales because of abnormally high values.

Two types of incentives were introduced to encourage participation and secure the quality and submission of the financial diary, (1) a diary incentive of 4 000 TZS for each diary filled out correctly and (2) a business prize competition. To qualify for the end of the year business prize of 1 mill TZS, a minimum of 7 out of 12 diaries had to be submitted. In addition, each participant was required to take part of monthly group meetings. Both formal and informal field visits and interviews were conducted throughout the year to observe the participants in their natural social environment.

Interviews and field visits

Ten months after the SL ended, we visited and interviewed 5 of the participants in their homes in Dar es Salaam. We made a ranked list of the participants based on their performance on each of the main indices, *HIKnow*, *HIMinds* and *HIAmb* (see Appendix C for their scores on each of the indices). The participants were chosen for visits based on this list. Three of them had some of the highest overall scores on the main indices, *HIKnow*, *HIMinds*, and *HIAmb*, and two had some of the lowest scores of our pool of participants. During the visits, we saw their premises and homes, shopped in their stores, met some of their family and got a peak into their society and neighbourhood. From the interviews, we

got insights into their SL journey before, during and after the project, and heard about their future plans, their thoughts about entrepreneurship and about the challenges they had to overcome. We use their stories to highlight some of our results.

From September to November 2014, we also met regularly with the SL leader and founder, Jacqueline Mgumia. Her insight was crucial to understand the data and mechanisms behind the financial diaries. Through the meetings, we received additional information about the characteristics of each participant, and their stories from the project year.

This paper mainly uses the data from the participant's financial diaries, but also makes use of the additional information from meetings with researcher Jacqueline Mgumia and insights from the field visits to five of the participants.

3.2.2 Data cleansing and reassuring data quality

The financial diaries were reported hand written on paper. Throughout the SL year, the data was plotted into EpiData. After the end of the project, it was exported to Excel, and subsequently crosschecked with the original paper entries. This is the point in time when we entered the project. We decided to import the data to Stata, both to be able to run quality checks on the data and for analytical purposes. Through this, we (1) corrected all the dates to match the calendar of 2013, (2) eliminated and corrected inconsistencies like strings or other symbols in the integer columns, (3) went through a significant share of missing values to make sure they were really missing, and (4) eliminated a number of instances of double reporting of weeks and days. Double reporting typically occurred at the transition of the month, as the financial diaries were reported weekly, and the month doesn't necessarily start on a Monday. As the participants solved this in different ways, it made accurate comparisons impossible. We rechecked all these instances to the original diaries, and manually corrected them in Excel before reimporting to Stata. Through this work we also got an understanding of the original data, and the collection method. Working alongside Mgumia, was a good way to learn about the participants, their business activities, and to understand their individual systematic reporting. An example of the latter is participant 46. For sales and stock, he would report the number of chickens he sold or had. A chicken cost about 4900 TZS, so we manually calculated the sales and stock by multiplying this by the number of chickens. We did equivalently with other instances like this one.

Where cleaning was not an option, we had to take certain measures with the data. If the participant had double reported weekly data, we made an average of the two entries. These were in total 125 instances across all the variables investments, savings, stock, loss and loans, and all the participants.

Subsequently, when all of these issues were solved, we graphed all daily and weekly data over the year for each participant (see appendix D for examples). We did this because we discovered some outliers, and also wanted to discuss the yearly development of the data for each of the 49 participants with Mgumia. This way we were able to either eliminate outliers, or get a sensible explanation for them. Participant 11 had reported stocks of over a 100 000, something Mgumia reacted upon immediately, so we corrected it for 10 000. Participant 2 is an outlier when it comes to sales, as his weekly average was over 80 000 above the next one. This data was, however, correct, as he is in the business of buying and selling cows, which are very expensive in Tanzania. Naturally, his sales would be high, despite no particular advantage in profits over the others.

There were entries handed over in December 2012 and January 2014. As the participants had to get into the routine and learn about financial diary reporting in December, and there were only a few reports from 2014, we decided to only focus on the year of 2013. Further, we also made the observations into full weeks and we started to analyse the data by week by including Monday, 31 December, 2012, and deleting Monday 30 and Tuesday 31 of December 2013.

3.2.3 Key field variables

Measuring business outcomes and entrepreneurial choices

There are five key field variables that we use throughout this paper, and they are divided into two groups. Firstly we have three key *successful outcome* variables, which are (1) business survival, (2) business sales and (3) profits, and will refer to these as simply the outcome variables. Secondly, we have two key *entrepreneurial choice* variables, which are (1) investment and (2) savings, and will refer to these as the choice variables. We sometimes in the continuation also include business purchases as a choice variable, to illustrate our findings on investment choices, as they are linked and it provides additional information.

Cowling suggests business survival as a success factor in itself (2007), and this is a reason why we use this as an indicator of competitive outcome. In line with (BBPT, 2014) we use sales and profits as measures of successful outcomes, and investments as a measure of competitive choices. In the research literature financial performance measures have been used to define small business performance. Some of these are tangible extrinsic outcomes such as revenue and revenue growth, profitability, number of employees (De La Paz Hernández Girón, León, & Domínguez Hernández, 2007; Haber & Reichel, 2005), financial performance, increased personal income and wealth (Paige & Littrell, 2002). However, financial measures are not always enough to measure success. For instance, subsistence businesses operate differently from large SMEs, so their view of success may therefore differ (Liedholm, 2002; Toledo-López, Díaz-Pichardo, Jiménez-Castañeda, & Sánchez-Medina, 2012).

In addition to investments, we use savings as a measurement of a positive field choice. The link between savings is seen as important to macroeconomic growth in underdeveloped countries (Deaton, 1989). On a country level, higher saving rates have been associated with higher income growth, and this has been seen as proof that savings leads to prosperity, and the lack of it leads to stagnation (Loayza, Schmidt-Hebbel, & Servén, 2000, p. 393). In addition to the macroeconomic benefits of savings, it is also important in a microeconomic perspective, as it evens out consumption and helps ensure the living standards for poor people facing volatile and unpredictable income (Deaton, 1989).

The frequency and detailed reporting enhances the probability of a dataset with less recalling problems and aggregation or calculation errors. This speaks in favour of trusting our data. For the reported hours of work in the business, we found days that had reports of up to 36 hours in a day. This has a cultural explanation of time perception, but it makes it hard to use for analytical purposes, and as a measurement of field choices. For reported loss and stock, there was not perfect consistency in how these were reported. Some reported the levels (e.g. current stock of clothes), and others reported for the current week (e.g. purchases of clothes). The financial diary also included a happiness index, which we were told they did not always report. The reported qualitative information was all in Kiswahili, and more related to anthropological topics, so we preferred our own English interviews of the participants, where we related the questions to more business and managerial inquiries.

We acknowledge that these variables could have been used to illustrate successful business outcomes and entrepreneurial field choices, but chose not to include them because of the above-mentioned reasons.

Aggregating data

To make the data analysable, we made some aggregations and adaptations based on the financial diaries, and based on personal interviews with Mgumia. In the following we explain the key field variables one by one.

Business survival

We generated a dummy variable for those who were perseverant and managed to have a business activity until the end of the project. Twenty-nine out of 49 had a business in the end.

Business sales

Unless stated otherwise, we make weekly averages of the monetary variables. We use the sum of the whole year per participant, and divide it by the weeks they have reported. We choose to average it in order to get as precise estimates as possible, as the daily observations contain noise, and some participants stopped in the middle of the year or did not submit all their financial diaries¹¹.

Profits

Because of high levels of miscalculations in the self-reported weekly profit variable, we decided to re-calculate it based on their daily reports of business incomes and expenditures. Equivalent to business sales, we subsequently did the same averaging procedure to the profit variable.

¹¹ To illustrate why this is a cleaner way than, for instance, creating totals, consider participant 28, who only submitted his diary 13 times (i.e. for 13 weeks). His total would then be considerably lower than many of the other participants who submitted the diary 52 times. By averaging, the total is divided by weeks, and we get more reasonable estimates.

Investments and business purchases

Investment costs include equipment (e.g. machines), site (furniture or renovation) and taxes. Investments are, for most, a one-time start-up cost for the participants. We therefore aggregate it to the total investment over the whole year instead of averaging it. To also get an estimate of their weekly investment levels, we generate an investment variable that includes the purchase costs of raw materials.

Savings

For savings the participants reported both on the amount and the type of savings (e.g. mobile savings, bank, relatives, etc.) the participants use. We summed these and also created a weekly average over the year per participant.

3.2.4 Description of the field performance

In the following we will describe the sectors and characteristics of the businesses of the 49 participants in our sample. We subsequently portray the participants' businesses and performance over the project year, 2013, based on the key field variables from the Social Lab, including business survival, sales, profits, investment and savings. We look at growth and development over the year.

Business sectors

The types of businesses that the pool of participants started during the project year can be categorized into six sectors: (1) Vegetable stands, (2) clothes and textiles, (3) poultry farming, (4) DVD/video shops, (5) food and drinks and (6) other.

Table 3-8: Business sectors

Sector	Freq	Percent
Vegetable stands	12	25.00
Clothes and textiles	14	29.17
Poultry farming	5	10.42
DVD/ video shops	5	10.42
Food and drinks	6	12.50
Other	6	12.50
Total	48	100.00

One participant got pregnant early in the project and didn't manage to start a business. She still took part in the project

Twelve out of 49 participants started a vegetable stand, 14 sold clothes on credit, shoes or produced textiles, five started poultry farming, five opened small video or gaming shops, and six sold food or drinks. The participants from group six did not fit into the mentioned sectors. In group six, one was an egg supplier and ran a bike transport business, one sold soap, one opened a training centre, one rented out a boat to fishermen and had a fishing shop, one had a small grocery store at home, and one was buying and selling cows. For the latter, the business activities were on a higher scale compared to the rest. Buying and selling cows is, naturally enough, a capital-demanding activity, and the sales and costs were at a completely different level than for the rest of the participants. Because of this, we will exclude participant number two in the following when we describe and analyse sales levels.

Business characteristics

All of the above are informal enterprises with no or only one employee. Their reasons for having an employee are often to be able to combine business with school, and the employee is usually a family member and thus not necessarily here to indicate business growth. Very few actually have business premises, and those with shops normally operate from home, and many sold their products on the street. The businesses are all small-scale commerce. Most of them are service-oriented and dependent on establishing a stable consumer base. This is, however, a challenge, as they typically operate in low-income areas, where the demand is limited and competition is fierce. Many of the participants had to change business sectors during the project year, or work extra jobs next to the main business activity due to low profitability in their business.

Performance

“Average weekly sales”: After adjusting for the largest outlier, the average business sales still have high variation (see Table 3-7), with a standard deviation of almost twice the size of the mean weekly value. The high variation may also be due to the fact that, for example, selling clothes gives higher sales than selling vegetables (see Appendix D), even though the end profit is not necessarily that different. See table 3-7 for the summary statistics of the variables that we use in this paper.

“Average weekly profit”:

High uncertainty and small margins in the business sectors can help explain the low weekly profits (see Table 3-7). Also here there is high variation, and it varies greatly between participants.

Some participants accumulated high levels of stock, but struggled to sell, due to lack of demand, while

others struggled due to hard competition or expiration of goods, etc. (clothes, vegetable stands). Another explanation for the low profit is inefficient management of the business.

“Total investment over the year”: 17 of the 49 participants have not reported any investment cost during the year of 2013. The types of businesses started by the participants typically required smaller purchases of stocks and goods, rather than bigger investments in equipment and registration costs. For those who did invest, the

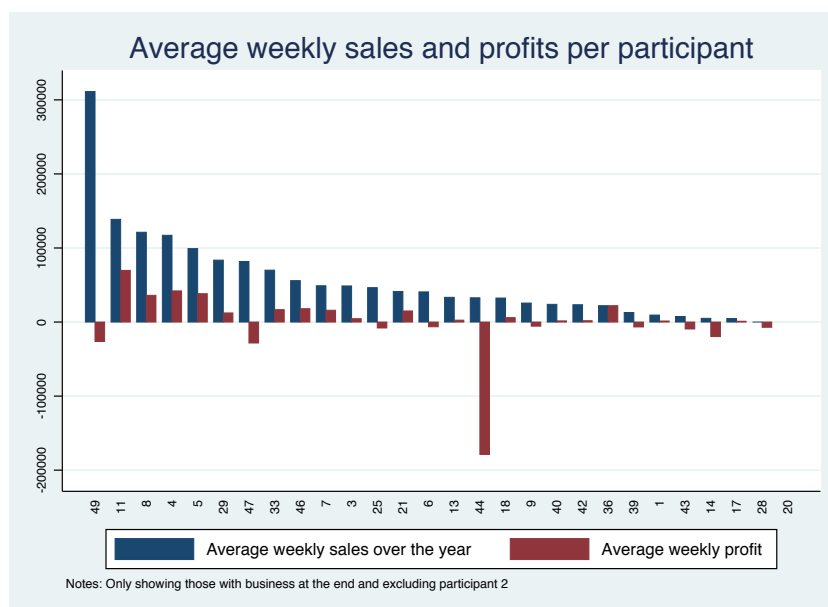


Figure 3-9: Sales and profit per participants from highest to lowest sales

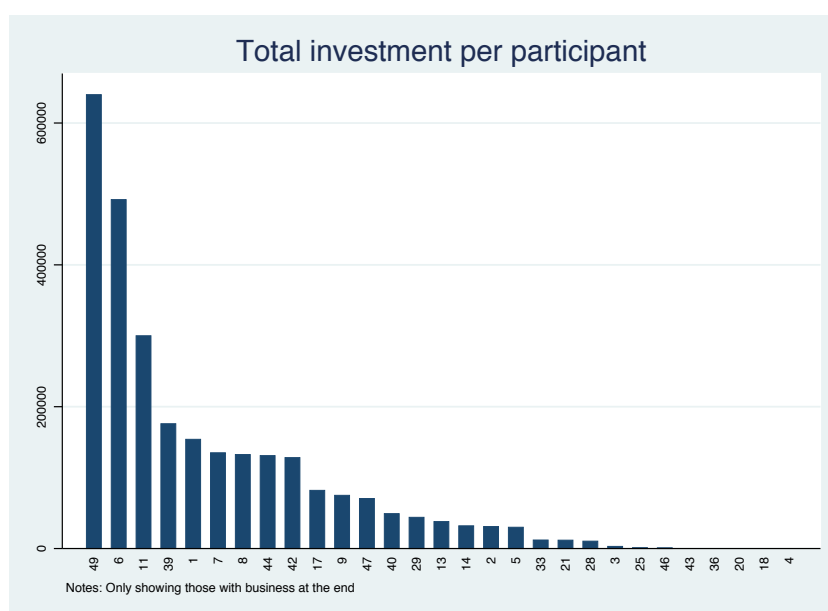


Figure 3-10: Total investment per participants from highest to lowest

investment cost was generally only reported at the beginning of the period¹².

“Average weekly savings”: The savings rate is quite similar to the average weekly sales, with high variations between the participants.

As the graph shows, a few participants drive the average savings rate up, and there is high variation between the participants.

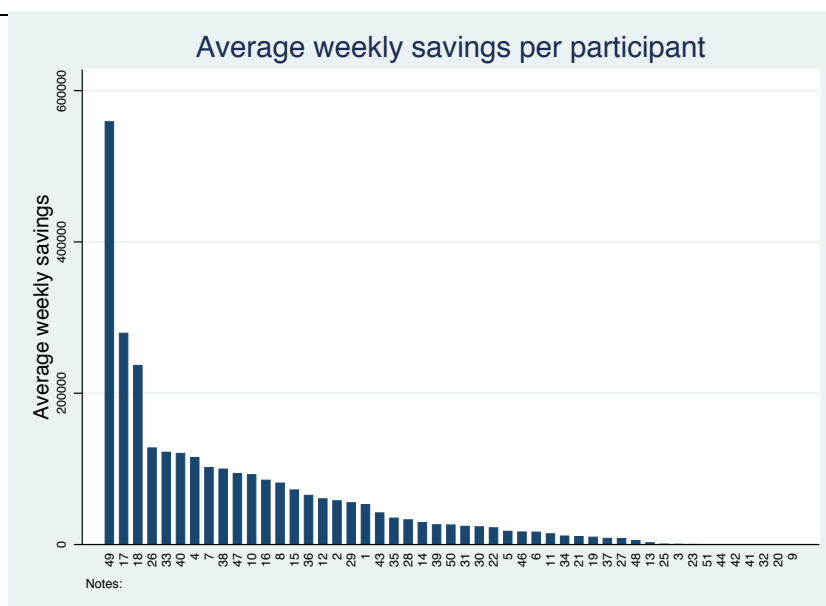


Figure 3-11: Average weekly savings from highest to lowest

Weekly development over the year

To get a measure of the growth in sales, profit and savings over the year for the whole sample, we did a panel data regression on the week variable, clustering on their ID numbers. We use a fixed effect over a random effect model since the uniqueness, or “the unobserved effect”, of the individuals is not likely to be uncorrelated with the independent variable, or a result of random variation. Another assumption is that the effect has to be relatively constant across time (Wooldridge, 2013). This (1) means that we suspect competitiveness and entrepreneurial abilities to be an unobserved effect that may affect sales, profit and savings. (2) It means we assume these unobserved effects for each of the participants is relatively constant over the year of 2013. We do this regression for descriptive purposes and to get an approximation of the size of the growth over the year. In this paper we will, however, not take a stand as to whether the other assumptions, like strict exogeneity, are likely to be fulfilled, as we do not take this analysis further.

¹² Looking at yearly development in investment is therefore not meaningful, and will be excluded in the growth description in the following section.

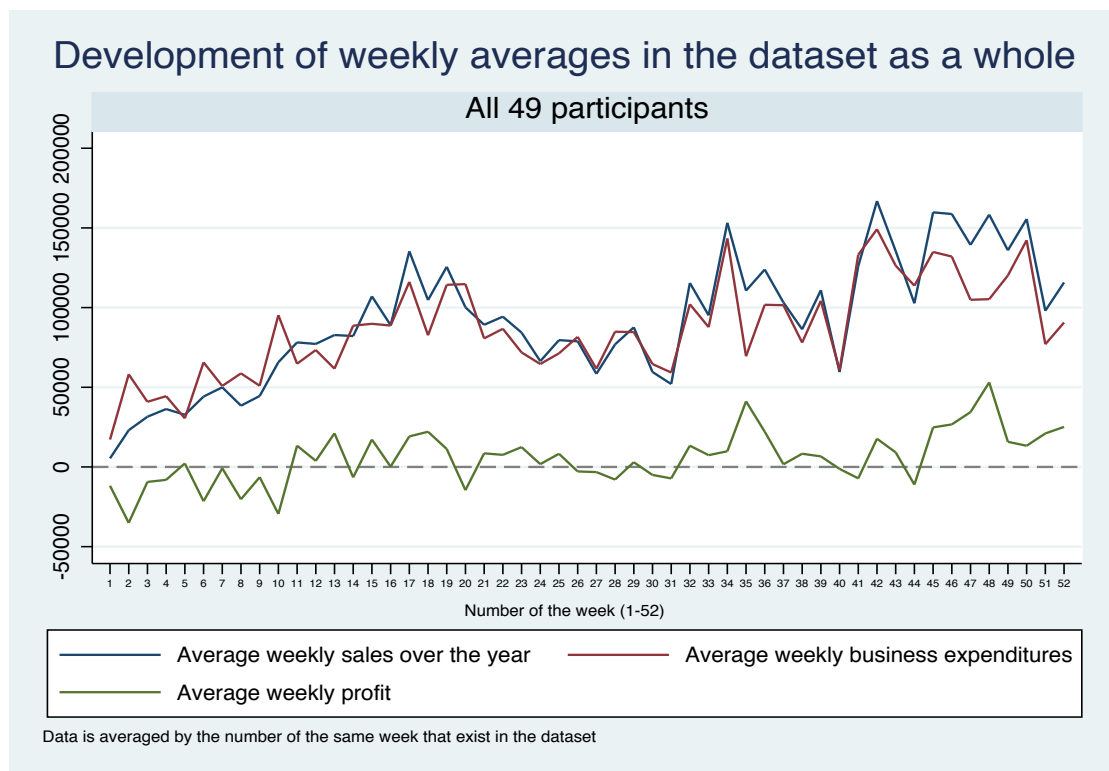


Figure 3-12: Development of weekly averages in the dataset as a whole. Sales and cost follow a similar pattern. The participants generally have high turnovers with small but frequent business purchases and sales, due to type of business e.g. vegetable stands with purchase up to twice a week. The trend is positive. Although the profit is low, it goes from a negative average in the beginning of the year to a positive average at the end of the year

The participants have high weekly growth in savings, especially during the last two quarters of the year. The high growth rate is somewhat driven by a few of the participants (participant 49).

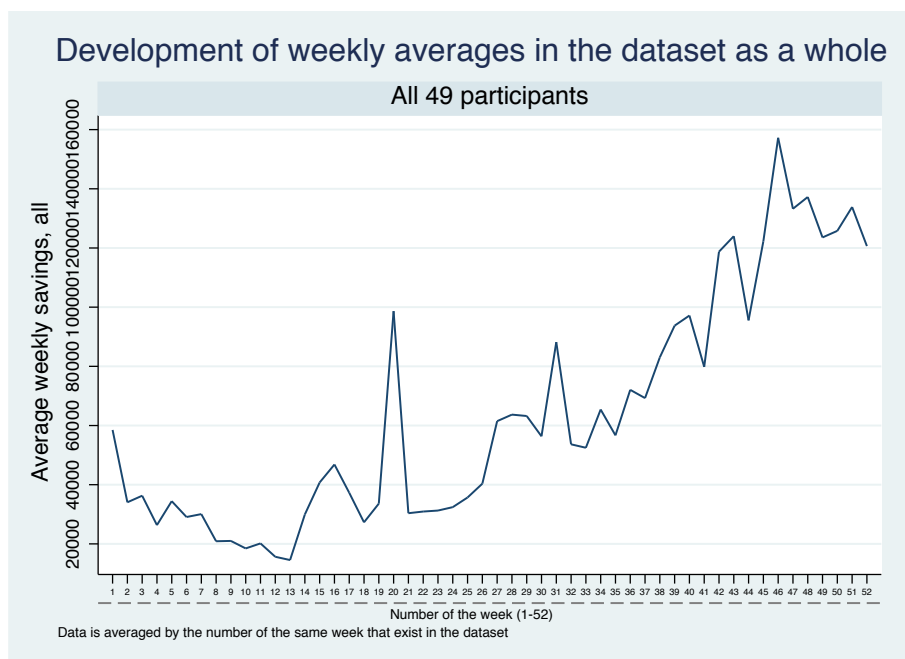


Figure 3-63: The development of average sales over the year

Table 3-9: Growth per week in TZS

TABLE: Growth per week in TZS

	Sales	Sales excl. participant 2	Profit	Purchases	Savings
Week nr	2193** (1068)	1355* (682)	651** (257)	1506* (861)	2031*** (737)
_cons	34391 (27168)	25525 (17334)	-11613* (6538)	39390* (21901)	9655 (18737)
R^2	0.036	0.033	0.013	0.024	0.052
N	2105	2054	2105	2105	2105

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In

Table 3-9, we see there is some significant growth in all four variables. A lot of the growth in sales is especially due to Participant 2, so we took him out to get a better picture of the remaining participants. The growth is still positive and significant at a 10% level. The growth is at about the equivalent of 1355 TZS per week. Profits and savings are significant at a 5% level, but savings is a lot higher. Business purchases also increase greatly over the year. All this seems to imply that they are increasingly growing their businesses. This development can also be seen by the trends shown in the graphs. All are positive, and without taking inference too far, this might suggest a positive effect of being in the Social Lab programme. To read more on this, see the upcoming PhD paper written by Mgumia.

4. Results

In the following chapter we present the two analytical methods used in this paper and the results. Part 4.1 is dedicated to a correlation analysis, and part 4.2 covers a regression analysis. Following, in 4.3 we present a discussion on gender differences. Concluding the chapter, we write some words about the other methods we considered for answering our research question, and about the limitations and validity of our findings.

4.1 Correlation analysis

To get an understanding of the associations between the lab and the field variables, we ran several correlation tests on the whole dataset. In section 4.1.1, we explain the chosen correlation method, before we present the correlation results in section 4.1.2 on the business outcomes, and in section 4.1.3 of the entrepreneurial choices. Section 4.1.4 presents the constraints that the young entrepreneurs face, and discusses business survival. This sets the ground for the following regressions in the next chapter.

4.1.1 Spearman correlation test

At first, we ran Pearson's correlations on all 49 participants taking out the most extreme outliers because the test is sensitive to nonlinearity and outliers in the data. However, since Pearson's correlation test assumes normality, and our data failed the normality tests (see Appendix A) we decided to instead use the non-parametric Spearman's (rho) rank correlation test¹³ to make the coefficients less sensitive to this. The coefficient measures the extent to which, as one variable increases, the other variable tends to increase without requiring that increase to be represented by a linear relationship (Statistics Solutions, 2014). The test works by first ranking the data, then calculating the Pearson correlation coefficient on the ranks. This means it ranges from -1 to 1. If the coefficient is at the extreme of -1 or

¹³ We also considered and ran Kendall's tau correlation test, which is another accepted non-parametric method to do a rank correlation. Spearman's rho usually has larger values than Kendall's Tau (Statistics Solutions, 2014). There is, however, no strong reason to prefer one test over the other, since they usually produce very similar results in significance testing (Colwell & Gillett, 1982). We therefore chose to focus on Spearman, because we have more experience with the test. As a robustness check we use Kendall's test results for comparison.

+1, the relationship is, respectively, *always* negative or positive. The null hypothesis is that there is no association between the variables (Gerald & Keller, 2012).

There are two assumptions behind the Spearman test. Firstly, the data is ordinal or continuous. Secondly, there is a monotonic relationship between the two variables tested (Lærd Statistics, 2014). Since it is not sensitive to outliers, like we have for some of the participants in our dataset, and since it also works for ordinal data, like some of our indices, we decided to use this test instead of the Pearson test.

We run the Spearman correlation test to investigate the relationship between the lab indices and the key field variables. We test to see which of the main indices, and mind-set indices tend to be more important for business outcomes and entrepreneurial choices.

First, we run correlations between the main indices and sub indices of *HIKnow*, *HIMinds* and *HIAmb*, against the business outcome variables, following a similar correlation test on the entrepreneurial choices variables. For all but business survival, we run the analysis on two levels:

- (1) For all 49 participants
- (2) For the 29 participants with business in the end

4.1.2 Business outcomes

The results from the correlation test on entrepreneurial outcomes are presented in Table 4-1. In the following we will discuss the results of each of the key outcome variables: business survival, sales and profits.

Table 4-1: Business outcomes correlation matrix

	Business Survival	Sales	Profit	Business Survival	Sales	Profit
HIKnow	-0,06	-0,17	-0,21	-	-0,02	-0,21
Business know	-0,15	-0,22	-0,29*	-	0,07	-0,31
Math grade	-0,03	0,04	0,06	-	0,08	0,05
Calculations	0,26**	-0,08	-0,20	-	-0,33*	-0,30
Entrepr.knowledge	-0,06	-0,16	-0,17	-	0,14	-0,16
HIMinds	-0,05	0,17	0,05	-	0,24	0,00
Patient	-0,05	0,20	0,07	-	0,32	0,15

Competitive	0,12	0,29**	0,19	-	0,41**	0,21
Risky	-0,13	-0,04	0,08	-	0,03	0,10
Confident	0,15	0,08	-0,01	-	-0,03	-0,16
HIAmb	0,04	-0,23	-0,38***	-	-0,24	-0,45**
Own business rank	-0,06	-0,16	-0,25**	-	-0,07	-0,14
Starting business	0,13	-0,16	-0,36**	-	-0,24	-0,47**
Training in entrepr.	-0,06	0,00	0,05	-	-0,05	0,00
Mentoring	0,17	0,27*	0,32**	-	0,26	0,41**
* 10% significance level ** 5% *** 1%	<i>All 49 participants</i>			<i>The 29 participants</i>		

Business survival

Except for calculation skills, there is almost a non-existent correlation between the indices and business survival. This would indicate that knowledge, mind-set and ambitions is not the restricting factors for whether a participant succeeds in keeping the business running until the end of the project year.

Sales

When we run correlations including all the 49 participants, *average sales* and the knowledge index do not seem to be correlated. Of the knowledge sub-indices, business knowledge is negative with a coefficient of -0.22, but it is not significant¹⁴. The results do not change much when correlating on the 29 participants with business in the end. However, we observe that calculation skills now have a negative correlation with high sales (-0.33), which is fairly significant¹⁵.

The mind-set index, *HIMinds*, does not seem to correlate with higher sales for all 49 participants. However, looking into the individual mind-set sub-indices, the correlations are stronger. *Patient* somewhat positively correlates with average sales (0.2). For *competitive*, the relation with sales is stronger and significant with a coefficient of 0.29. Neither *risk* nor *confidence*, seem to have any impact on business sales. For the 29 participants correlation, we now find a positive correlation of the mind-set index, *HIMinds*, (0.24), and the

¹⁴ When we refer to a non-significant result, use a significance level of 10%.

¹⁵ Throughout the analysis, a fairly significant result refers to a significance level of 10%, a significant result refers to a significance of 5%, while a strongly significant result refers to a significance level of 1%.

correlation coefficient of *patient* increase from 0.2 to 0.32, though it is still not significant. More importantly, *competitive* is now strongly and significantly correlated with sales with a coefficient of 0.41.

As for the ambition index, *HIAmb*, on all 49 participants the correlation coefficient is -0.23, and is thus somewhat negatively, but not significantly, correlated with sales. The relation does not seem to change when we only look at the 29 with business (-0.24). It seems like higher overall ambitions relate to lower levels of average sales for both samples.

Profits

The relation between the knowledge index, *HIKnow*, and *average profit*, is somewhat negative, but not significant looking at all 49 participants. At the sub-index level, business knowledge is both fairly significant and negatively correlated with average profit (-0.29). The correlations stay more or less the same when we only focus on the 29 participants. High overall knowledge level (*HIKnow*) and specially business knowledge indicate lower average profit.

For the mind-set index (*HIMinds*) the correlation for the 49 participants is as it was with average sales, almost non-existent. However, in this case, none of the sub-variables are correlated either. This can be due to less error in the sales reporting compared to the much more detailed cost reporting in the financial diaries. When we only look at the 29 participants with business until the end, the correlation stays weak for *HIMinds*. However, *competitive* is now positively, but not significantly, correlated with average profits (0.21).

For the ambition index, *HIAmb*, the picture is quite different. The index is negative and, strongly significantly correlated with an average profit with a correlation coefficient of -0.38 looking at all 49 participants. For the 29 participants, the correlation coefficients increase and continue to be significant. The absolute value of some of the coefficients increase; *HIAmb* increases to -0.45, and we notice that the sub-indices *starting business* increases to -0.47 and *mentoring* to 0.41, and the latter coefficient is positive. Both are highly significant.

Summary of business outcomes

We find that *competitive* positively correlates with average sales, and is significant both for all 49 and for the 29 participants. It also somewhat correlates with average profit. For *HIAmb*, we find strong and significant negative correlation to average profit, and even more

so for the 29 surviving businesses, which might seem like the ambitious ones stay running their businesses longer even though it is unprofitable.

4.1.3 Entrepreneurial choices

The results from the correlation test on entrepreneurial choices, is presented in Table 4-2. We will in the following discuss the results on each of the key choice variables: investments, purchases and savings.

Table 4-2: Entrepreneurial choices correlation matrix

	Investments	Purchases	Savings	Investments	Purchases	Savings
HIKnow	-0,23	-0,13	0,30**	-0,18	0,11	0,41**
Business know	-0,29*	-0,16	0,30**	-0,18	0,24	0,44**
Math grade	-0,07	0,01	0,12	-0,02	0,10	0,25
Calculations	0,11	-0,02	0,18	0,05	-0,19	0,16
Entrepr.knowledge	-0,32**	-0,15	0,33**	-0,22	0,25	0,41**
HIMinds	0,23	0,21	-0,31**	0,27	0,27	-0,41**
Patient	0,15	0,23	-0,02	0,11	0,37*	-0,05
Competitive	0,28*	0,29*	-0,17	0,25	0,37*	-0,07
Risky	0,13	-0,07	-0,03	0,17	-0,08	-0,25
Confident	-0,15	0,15	-0,22	-0,09	0,10	-0,24
HIAmb	0,20	-0,10	0,13	0,29	-0,03	0,08
Own business rank	0,23	-0,06	0,14	0,36*	0,00	0,08
Starting business	0,07	-0,04	0,14	-0,04	-0,08	0,02
Training in entrepr.	0,09	-0,08	-0,18	-0,01	-0,02	-0,03
Mentoring	-0,03	0,16	-0,11	-0,24	0,07	-0,15
* 10% significance level ** 5% *** 1%	<i>All 49 participants</i>			<i>The 29 participants</i>		

Investments

For all 49 participants, *HIKnow* correlates negatively with investments, but not at a significant level. At the sub-index level, *business knowledge* and *entrepreneurship knowledge* are also negative, and significant on a 10% level and 5% level respectively. When only looking at the 29 participants, the knowledge correlations decrease and none are significant.

HIMinds is positively correlated with investment (0.23). However, *competitive* also stands out as the strongest relation and as the only fairly significant of all the mind-set indices, including the main index. As with the knowledge index, this relation somewhat decreases from 0.28 to 0.25 when we only look at the 29 with businesses.

Also for *HIAmb*, the correlation is somewhat positive, but not significant. The sub-index “rank of having own business” is positively correlated with *investment*. The *HIAmb* correlation coefficient increases from 0.20 to 0.29 when we only look at the 29 participants. We can observe the same increase for “rank of having own business” that goes from 0.23 to 0.36, the latter being fairly significant.

Business Purchases

In the case of *business purchases*, the knowledge index, *HIKnow*, does not seem to have any correlation, neither for the whole sample nor when only looking at the 29 with business until the end. It is only at the 29 participants’ levels where we find some correlation when run against the sub-indices *business knowledge* (0.24) and *entrepreneurship knowledge* 0.25. However, none of the two are significant.

We do find similar correlation with *HIMinds* and business purchases as we saw with *investment* on both levels. Also in this case, the correlation coefficient increases from 0.21 to 0.27, but is not significant. As with the previous field variables, it is *competitive* of the mind-set indices that has the strongest correlation with business purchases, and it increases considerably from 0.29 to 0.37 and both are fairly significant. *Patient* also correlates positively with purchases and like *competitive*; it increases from 0.23 to 0.37 (the latter is significant at 10% level).

Neither the ambition index (*HIAmb*), nor any of its sub-indices seem to have any correlations with business purchases.

Savings

The correlation is strong and significant between *HIKnow* and *average savings* on all levels, with the correlation coefficient of 0.30 when using all 49, and 0.41 when only using the 29 participants. It is particularly *business knowledge* and *entrepreneurial knowledge* of the knowledge sub-indices that correlate with savings, also here on both levels. The correlation coefficient for *business knowledge* is positive and significant, and increases from 0.30 to

0.44, while for *entrepreneurial knowledge* the coefficient is significant and increases from 0.33 to 0.41.

We also find a strong, significant, but negative correlation between the mind-set index (*HIMinds*) and average savings (-0.31). Of the mind-set sub-indices, only *confident* is somewhat negatively, but not significantly correlated. The correlation coefficient's absolute value increases when we take out the 20 participants without a business (-0.41). Now both *confident* and *risky* becomes somewhat negatively correlated (approx. -0.25). None of the mind-set sub-indices coefficients are significant.

Also here, neither the ambition index (*HIAmb*), nor the sub-indices seem to have any correlations with *average savings*.

Summary of entrepreneurial choices¹⁶

None of the main indices correlates more with investments than the others, and only some of the sub indices have significant correlation, though not on both levels. Again it is the mind-set index, *competitive*, that is the strongest of the mind-set variables, here against investments and purchases.

Overall we see a clear tendency of the knowledge indices to positively correlate with higher average savings. The mind-set index (*HIMinds*) is negatively correlated with this key field variable, which is mostly driven by *risky* and *confident*.

Main findings from the correlation analysis

In the correlation analysis, *competitive* is the entrepreneurial characteristic from the lab that most consistently correlates with successful outcomes and entrepreneurial choices^{17,18}.

¹⁶ We also ran Pearson correlation and Kendall's rank correlation tests on our variables. Spearman and Kendall's tests show a different kind of association, namely on ranks, not linear correlation or dependence, as does the Pearson test. Spearman's results are very similar to the results from Pearson's test. We find the knowledge index to be strongly positively correlated with average savings, and *competitive* correlates positively with profits, sales and investment. Kendall's test also goes in the same direction, although this test does not show coefficients as strong. This test usually has lower coefficients than the Spearman test.

¹⁷ For robustness checks, we also ran partial correlations on *competitive* against sales and investments, controlling for each of the variables for math grade, confidence and willingness to take risk. Partial correlation measures the degree of association between two variables, and has the additional feature of removing the effects of other variables (Stata, 2014). *Competitive* is still strong and significant. See description of how we did the test and the details about the results in Appendix E.1.

We find that high knowledge is the most important index for higher savings. Human capital is, in this paper, represented in the form of high knowledge (*HIKnow*), and is by many considered one of the crucial traits for business survival and growth (Liedholm, 2002). However, for our pool of participants, this trait does not seem to be of as much importance to achieve successful business outcome, as for savings. We also find strong negative correlation between *HIAmb* and profit.

4.1.4 Constraints and business survival

Only 29 of the participants managed to run their businesses until the end. However, this does not mean that the remaining 20 lacked the necessary knowledge, mind-set and ambition needed to run a business. Business survival was not strongly correlated with these aspects. Theory however says that they are important to succeed in business. In our dataset, we found that the participants could possess these skills and mind-set, but that other constraints kept them from running the business until the end.

Table 4-3: Reasons why 20 participants dropped out of the SL Project

Reason for drop-out	Number of participants
Pregnancy	4
Family issues	4
School	4
Waged work	3
Theft	1
Other	4
Total	20

Some of the 20 quit their business despite the fact they managed to run their business well. As we can see in Table 4-3, some of them went back to school, or got a waged job somewhere else. Self-employment was, in this case, a stepping-stone to further or formal employment. Furthermore, three of the smartest girls in our dataset either got pregnant or had to start taking care of family members. Some had other family issues, and they could not run a business alongside this. One was subject to theft in his business in his home, and did

¹⁸ Since *competitive* is a binary variable, we also ran a point biserial correlation test, which is a useful test when we want to correlate a binary variable with a ratio variable (Dolgun, 2012). *Competitive* is still strong and significant. See Appendix E.2.

not manage to get back into business. As we can see, knowledge, entrepreneurial mind-set and ambitions were not necessarily the constraining factors for many of the participants in our pool. Actually, when we ranked and investigated the participants based on their scores on *HIKnow*, *HIMinds* and *HIAmb*, we find the participants with the surviving businesses spread out on the ranking lists.

On the other side, we have the 29 that survived. Some still managed through, despite facing challenges like theft, tax and registration problems, police, and financial constraints. One was subject to theft three times, but was reimbursed by the family. The tax authorities visited one and confiscated all his stock, as he had not formally registered his business. He had sufficient savings to pay a fine and start another business. Participant 2 was, at one point, in possession of stolen cows, and was arrested by the police. He, however, managed through, and went back into the same business.

Some of the participants with surviving businesses received substantial financial support from family or income sources other than the business itself. Both financial and human capital is an important element in a start-up face of a new enterprise. Our pool of participants is, in general, a group faced with high capital constraints, often too small to consider taking up microcredit loans. Access to other financial sources could, therefore, be very important for survival through the start-up face and dealing with high uncertainty. The reasons why some businesses survived and some stopped might go beyond these aspects of knowledge, mind-set and ambition needed to run a business. However, for the other outcome variables, and also the choice variables, we find significant impacts of these traits. We, therefore, focus our continued analysis on the 29 participants that survived. We consequently also exclude business survival from the outcome variables, as we have shown that there are other things constraining this than the aspects that the indices capture. The regressions including all 49 observations and on business survival as a dependent variable, can however be found in Appendix F, Table F.1. In BBPT et al. (2014) the sample of interest consists of adult and established entrepreneurs. Taking out those who did not really succeed in running a business makes the data a better fit for comparison, and more meaningful to study the levels of business outcomes.

4.2 Regression analysis

Despite the fact that not all the usual assumptions for estimating our model by Ordinary Least Squared (OLS) are present, the model can still be informative and useful. Without worrying about the underlying population model, and the usual properties of efficiency and unbiasedness, we can still use the model for predictions (Wooldridge, 2013) of the key field variables for a set of young Tanzanian entrepreneurs with similar characteristics to our pool of participants. The model also contributes descriptively. The correlations revealed certain correlations in the data. These results are more robust, as we use a non-parametric method. The impact of *competitive* is especially robust, as we have run several different tests. However, now we are interested in understanding the magnitude of the impact of the indices on the key field indices, i.e. the monetary change that the indices are associated with. These cannot be conclusive, but indicative and descriptive of our pool of participants.

We will start running regressions on business outcomes (Table 4-4) then on entrepreneurial choices (Table 4-5) on both level and log form (for the latter, see Appendix F, Table F.3 and Table F.4). The reason why we log the data is to scale down some of the outliers in our dataset, and to get a percentage display of the impacts.

The structure and the naming convention of the regression models are as follows. We run three models for each key field variable. For the a-models, we run the key field variables against the mind-set sub-indices, since we are especially interested in the effect of the mind-set sub-indices. For the b-models we introduce the *HIKknow* and the *HIAmb* indices. For the c-models we additionally include the control variables. The numbers of the regressions on the *outcome* variables (excl. business survival) are respectively (1) for sales, and (2) for profit. Then for the *choice* variables, the numbers are (3) for investments, (4) for purchases and lastly (5) for savings.

In the b-models we only include the main indices for *HIKknow* and *HIAmb* due to the problem of multicollinearity, since the sub-indices in each of them are highly correlated with each other¹⁹. We are also more interested in seeing the effect of these seen

¹⁹ When correlating the *HIMinds* sub-indices, we do find not strong correlations. Additionally, when we run Spearman correlations of *competitive* directly on these four variables for the whole dataset, we only find *confident* to have a correlation with *competitive* at a coefficient over 0.2 (absolute value). Confidence thus seems, interestingly enough, more influential than actual mathematical skills when choosing to compete.

together, and to use them to control the effect on the mind-set indices, to make it more likely for the zero conditional mean to hold. For this same reason, we additionally introduce background information (controls) in the c-models, to control for other factors that may affect the dependent variables and be correlated to the independent variables. We will, however, still not be able to infer causality due to other unfulfilled assumptions. Again due to multicollinearity, we do not include the aggregated *HIMinds* index or age as a control in the c-models, as it is highly correlated with gender. There are no females over 22 years in our dataset.

As we have stated, *competitive* is a variable of extra interest to us, and it also seems to be one of the most relevant. Since competitiveness was measured by asking if the participants wanted to compete in mathematical calculations, one would assume that both mathematical skills (math grade and calculation skills) and confidence in how well they had done on the calculations compared to the others (*confident*) and could affect this decision. “Willingness to take risk” is also likely to have an impact on this choice. In the correlations section, we ran partial correlations to control for this, without any change in the result. Now in the regressions we include them in all the models to control for their partial effect in explaining the variation in the dependent variable.

All the monetary measures in these regressions are in 1000 TZS, unless specified otherwise. We take Participant 2 out of the regressions on sales, as his sales level is abnormally high.

4.2.1 Business outcomes

The results of the regressions on the key choice variables are found in Table 4-4. Following, we describe the models. The dependent variable is specified in the heading.

Sales

The patient participants throughout have higher sales at about 13 TZS per week for each increment of patience level. The competitive participants stand out selling significantly more than the non-competitive at 67-72 TZS per week. The more risky and confident participants are the less they sell, also when introducing *HIKknow* and *HIAmb*, and the controls. *HIKknow*

and *HIAmb* have a positive effect on sales. Only *competitive* is significant in the three models.

When logging the dependent variable (see Appendix F, Table F.3) the coefficient of *patient* is significant at a 10% level, and associated with an approximate increase in sales of 50%. This stays when introducing *HIKnow* and *HIAmb*, but loses its significance when introducing the control variables. The coefficient of *competitive* is significant throughout and associated with an approximate increase in sales of 250-270% above the non-competitive. The coefficient of *confident* is significant at a 10% level in the first regression, but loses its significance in the next two. It is associated with an approximate decrease in sales of 30% for each level of *confident*.

Profits

We see that the more patient a participant is, the more profit he is left with. The coefficient increases from 7-9 TZS when introducing *HIKnow* and *HIAmb* and then the control variables. However, it is not significant. *Competitive* has a significant impact on profits when we run the regression on the mind-set sub-indices. This indicates that the competitive participants, on average, generate a profit of 32 TZS more per week than the non-competitive. This significance is, however, neutralized, and the impact reduced when introducing the *HIKnow* and *HIAmb* indices, which both seem to have a negative effect on profits, although only the latter is significant. Introducing controls does not have a mentionable impact on the coefficients, but reduces the significance of *HIAmb* to a 10% level.

Summary of business outcomes

Generally, we (1) find that the competitive participants sell more than the non-competitive, and manage to keep the profits positive. This variable is the most consistently significant among the mind-set sub-indices, (2) that the patient participants sell less, but are left with more profit, and both risky and confident participants generally sell more, but don't remain with as much, as they tend to have lower profits (3) that raise scores on the indices *HIKnow* and *HIAmb*, and are associated with lower sales and profits.

Table 4-4: Regression results of business outcomes on indices. 29 participants

	Sales (1a)	(1b)	(1c)	Profits (2a)	(2b)	(2c)
Patient	13.498 (9.481)	12.979 (9.626)	12.620 (10.366)	7.463 (5.766)	9.092 (5.482)	9.132 (5.632)
Competitive	67.036** (30.842)	72.917** (33.262)	67.430* (36.775)	32.469* (18.622)	24.035 (18.188)	18.395 (18.869)
Risky	-8.371 (19.563)	-6.067 (20.311)	-6.278 (22.891)	-8.377 (12.494)	-10.772 (11.947)	-8.292 (12.044)
Confident	-3.504 (5.394)	-4.017 (5.515)	-2.834 (5.950)	-5.503 (3.359)	-5.659* (3.224)	-4.744 (3.284)
HIKnow		28.481 (50.431)	10.127 (63.574)		-12.249 (31.798)	-28.480 (35.402)
HIAmb		11.834 (12.844)	15.789 (14.561)		-17.451** (8.026)	-15.640* (8.331)
Female			-1.982 (29.404)			-18.603 (16.065)
Business_stream			-17.337 (28.461)			20.404 (15.794)
Newspaper			-16.842 (28.165)			-4.860 (16.444)
Own_bus			32.540 (27.601)			20.964 (15.612)
Eat_meat			-7.642 (14.684)			-8.175 (7.565)
_cons	22.903 (50.007)	-31.560 (72.232)	3.986 (90.943)	14.937 (33.962)	69.491 (45.687)	87.775 (53.501)
R^2	0.213	0.261	0.363	0.164	0.324	0.494
N	28	28	28	29	29	29

The regression on sales is excluding participant 2, because of abnormally high values. The rest include all 29 participants.

Sales and profits are measured in 1000 TZS

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Std. errors in parenthesis

4.2.2 Entrepreneurial choices

The results of the regressions on the key choice variables are found in Table 4-5. In the following, we describe the models. The dependent variable is again specified in the heading.

Investments

We find that patient people invest slightly more, but the effect almost disappears when we introduce the other indices and the control variables. Competitive participants invest, on average, 128 to 164 TZS more than non-competitive participants. This is a huge increase in comparison to the constant and is five times the constant in the first model and 20 times the constant in the third model. The competitive coefficient is significant at a 10% level and the significance increases in the two following models. The riskier participants invest, on

average, between 20 to 30 TZS more for each risk level, but this is not significant. The coefficient of *confident* is negative throughout, and around 2 to 3 TZS. *HIKnow* goes from being positive to negative. The variation is very big, and we do not see much effect of knowledge. The ambitious participants invest some more than the non-ambitious.

When logging the dependent variable, we find that the coefficient and the significance level of *competitive* increase over the models. It is associated with an increase in investments of 200 to 300% above the non-competitive (see Appendix F, Table F.4).

Business purchases

When we run the regression on business purchases, we find more significant results on mind-set indices than we did for investments. All coefficients also increase considerably, which means that the pool of participants on average spent more on business purchases per week than they did on investments for a whole year. Patient participants generally spend significantly more, while risky and confident participants spend less. *Competitive* stands out with a coefficient ranging around 250 TZS, and highly significant in the a-and b-models and also significant in the c-model.

Also in the logged version of the models, we find *competitive* to stand out with a significant increase in spending of about 200 to 300% on business purchases.

Savings

The mind-set indices do not have a significant impact on savings, but the patient and competitive generally save more. The coefficient of risky is ambiguous. The confident generally save less. *HIKnow* seems to be the most determining index for savings levels, and the participants with higher knowledge save on average 126 TZS more for each unit increase in knowledge level when the other variables are controlled for. This is significant.

In the logged version we find the risky and the confident to save significantly less. *HIKnow* is still significant in the b-model and decreases to a 10% level in the c-model. The participant saves around 300-400% more for each unit increase in score on the *HIKnow* index.

Table 4-5: Regression results of entrepreneurial choices on indices. 29 participants

	Investments			Purchases			Savings		
	(3a)	(3b)	(3c)	(4a)	(4b)	(4c)	(5a)	(5b)	(5c)
Patient	4.117 (20.018)	0.045 (20.217)	0.436 (22.440)	70.966** (26.918)	64.399** (26.931)	45.586* (26.152)	7.148 (17.025)	5.709 (14.107)	5.749 (9.940)
Competitive	128.564* (64.649)	150.242** (67.073)	163.880** (75.186)	249.583*** (86.934)	286.605*** (89.350)	238.287** (87.622)	39.847 (54.985)	40.402 (46.804)	38.265 (30.912)
Risky	27.128 (43.374)	31.522 (44.058)	20.140 (47.994)	-105.840* (58.326)	-104.296* (58.691)	-105.417* (55.932)	-16.405 (36.891)	4.273 (30.744)	4.288 (18.958)
Confident	-3.176 (11.660)	-2.413 (11.889)	-2.036 (13.087)	-31.521* (15.679)	-28.997* (15.838)	-19.542 (15.252)	-6.180 (9.917)	-10.371 (8.296)	-5.266 (5.704)
HIKknow		9.648 (117.266)	-22.853 (141.069)		-57.374 (156.214)	-187.247 (164.402)		255.065*** (81.829)	126.289** (57.709)
HIAmb		42.148 (29.600)	51.120 (33.195)		62.830 (39.432)	67.506* (38.686)		32.660 (20.655)	22.624 (13.499)
Female			10.264 (64.017)			-99.745 (74.605)			-33.470 (26.784)
Business_stream			-81.958 (62.935)			70.845 (73.344)			-27.567 (30.112)
Newspaper			-56.025 (65.523)			23.431 (76.361)			-48.898* (27.260)
Own_bus			7.678 (62.208)			94.384 (72.498)			26.475 (26.839)
Eat_meat			11.975 (30.144)			77.512** (35.129)			1.133 (14.054)
_cons	25.610 (117.908)	-88.132 (168.484)	8.599 (213.189)	111.970 (158.553)	7.253 (224.445)	-65.972 (248.450)	101.025 (100.283)	-210.837* (117.570)	-32.590 (92.227)
R ²	0.195	0.265	0.358	0.357	0.424	0.615	0.028	0.403	0.339
N	29	29	29	29	29	29	29	29	49

Std. err. in parenthesis

Investments, purchases and savings are measured in 1000 TZS

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Summary of entrepreneurial choices

Generally, in our sample, we (1) find that the competitive participants invest a lot more than the non-competitive, and this finding is even stronger for business purchases. This variable is the most consistently significant among the mind-set sub-indices. Further, we (2) find that the patient participants do not invest more, but they have higher business purchases, and also save more; (3) that risky participants invest more, but spend less on the day-to-day business purchases; (4) the confident generally both invest, have business purchases and save less; and (5) that higher score on the index *HIKknow*, is strongly associated with higher savings, and on *HIAmb* generally all coefficients are positive.

4.2.3 Findings from the regression analysis

The purpose of running regressions was to highlight the findings from the correlation analysis, and give the associations that we found a more descriptive and tangible interpretation.

As we saw from the correlations, there are four strong relations discovered in our dataset, namely how competitiveness has a strong and significant association to (1) average business

sales, (2) total investments over the year, (3) how *HIKknow* is positively associated to average weekly savings, and (4) how *HIAmb* is negatively associated with average profit. We will draw these findings and connect them to the regression analysis

We see that *competitive* is not only correlated, but is associated with an increase in sales and investments of great magnitude, which is double or triple the results of the non-competitive. This association is in line with the findings of other researchers. Zhang (2013) and Buser et al. (2014) also find competitiveness to be decisive for performance. BBPT (2014) find competitive entrepreneurs to have higher sales and profits, although they find stronger evidence on profits than for sales. The reason why we find stronger relations to sales than to profits might have to do with the fact that they investigated already established entrepreneurs, while our pool is in the start-up year. Furthermore, BBPT (2014) find a strong and robust association of competitiveness on investment choices. Our findings confirm this. The strong association we find with business purchases further underpin this finding.

Higher knowledge is associated with a lot higher savings. For each unit extra of *HIKknow*, they generally save up to 200-300% more, or 100-200 more per week. The sub-index *patient* also seems to relate positively to average sales. As previously argued, savings contributes positively to both in a micro- and a macroeconomic perspective. We also find advantageous impacts of patience, in line with Stutter et al. (2013). We, however, did not find patience to predict a higher savings rate, like they do, but this might be due to our small sample size.

The more ambitious the participants in our dataset tend to have lower profit. This might indicate that they are more persistent, even though the business does not go very well. As it usually takes time for a business to turn profitable, and we only have data for the start-up year, we do not take the analysis of this association any further.

To sum up, the strongest findings from the regression entrepreneurial choices are large and significant impacts of (1) *competitive* on investment and business purchases both on level and log form and of (2) *HIKknow* on savings.

4.3 Gender differences

The final price of one million Tanzanian shillings to the participant that managed to run the business best created a competitive environment during the project year. On our field visits,

we got the impression that both male and female participants enjoyed this environment very much, and almost all of the 49 participants expressed that they believed they would be the winner of the competition. In the following section we will therefore take a deeper look into the gender differences on performance in a competitive environment.

During the RJ lab, 33% of the male participants and 21.4% of the females in our pool of

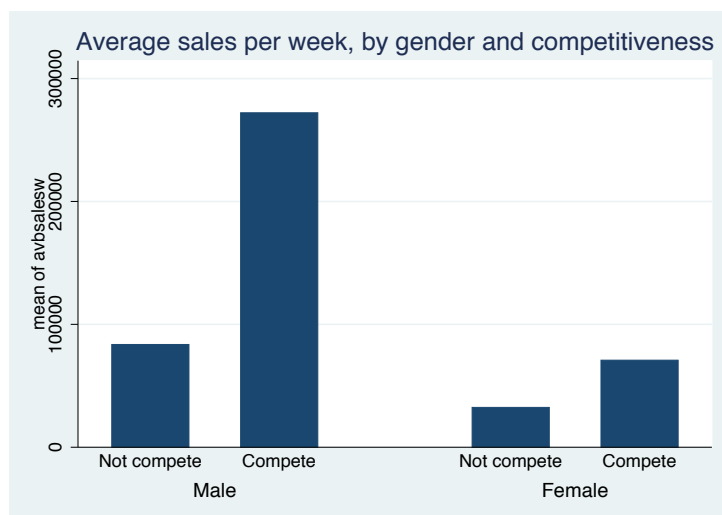


Figure 4-1: Competitiveness and gender differences in sales performance

participants chose to compete.

When running a proportion test, we find that this difference is not significant, nor does the regressions show any significant gender effects on the key field variables for the 29 that manage to have a business until the end. The gender difference becomes stronger when we run regressions on all 49 participants.

Now the effect is significant on

business survival, business purchases and highly significant for the logged version of savings (see Appendix F; Table F.1 and F.2). Due to lack of normality, we supplement the regressions with Wilcoxon rank sum tests on the key field variables and gender. However, none of the tests are significant²⁰.

Our findings are in line with that of Zhang (2013); competitiveness relates positively to performance in the field, but the gender effect on competition outcomes and choices is not as clear based on the statistics. Despite lack of significance, we do observe some tendencies of lower performance in the key field variables for the female participants compared to the male. This is illustrated graphically with average sales, gender and competitiveness for all 49 participants. The graph shows a difference between males that compete and sales level, and the rest for all 49 participants. The difference between competitive women is not as clear, nor is the difference between competitive females and non-competitive males.

²⁰ We also ran a Wilcoxon rank sum test between competitive and non-competitive female on key field variables, but we did not get significant results. The lack of significance in the Rank sum test might be due to low number of observations.

This is supported by experimental measurement on competitiveness like Niederle and Vesterlund (2007) and Gneezy et al. (2003). In our pool of participants we observe a tendency of male participants to outperform women in the competitive environment the SL was conducted. We also see, however not significantly, that fewer females enter the competition relative to their male participants.

4.4 Limitations and validity

As always when dealing with self reported data, there is a danger of non-compliance to the provided rules of reporting. For instance, we acknowledge that in small scale business reporting, the entrepreneurs sometimes don't have a clear distinction between personal and business economy. In our case, the interpretation of savings is thus not only reflecting business. Additionally, participant 20 would report all his income as personal, since he ran a training centre and had no sales in the usual sense. As we were not in the position to distinguish between his personal income and income from the training centre, we provided him with the lowest observation of sales. We had to take such instances into account when drawing conclusions from our data.

The RJ experiment specifically aimed at and measured the effect on knowledge, mind-set and ambition. This was good for our analysis, as we needed a measurement of these dimensions. However, since we have members of the treatment group of the RJP training in our pool of participants, these are likely to now have a higher score on these traits, than in the baseline, and than the average Tanzanian youth in Dar es Salaam. Also the two groups that received training through the UDEC program, are likely to know more about entrepreneurship now than in the RJP baseline, and more than the average Tanzanian youth. This makes our pool less representative of the population. Our results can however still be valuable given a similar sample of youth, and the relations within a similar pool of entrepreneurs can still apply when this is taken into account.

One might argue that the measurement of *competitive* is weak, or at least weaker than the other mind-set measurements, as it is based on only one question. However, in the paper of BBPT (2014), they asked this question several times. They did not get different results when regressing on the choice of competing one time versus several times.

As for the regression models, we know that the assumptions behind OLS like linearity in data, homoscedasticity and the zero conditional mean, are not fully complied with. We acknowledge that the models might not be unbiased and efficient. They, however, give useful insights into the dynamics of our data, and can indicate interesting relations to be investigated in future analysis. We, however, decided to run *robust* regressions, because of some outliers in our dataset, our three main findings to see what would happen to our results. Also, the fact that the participants met in groups and have different treatment status, might have affected the assumption of independence between the observations. We therefore also ran correlations *clustering* on group level. This naturally decreased the significance levels of the models. For each of the three relations found in the dataset, we thus provide how the significance changed: (1) Sales-compete relation: When running a robust regression on sales, *competitive* is still fairly significant in the a- and b-models, but not significant in the c-model. Clustering eliminates the significance of *competitive*, but it is still fairly significant for profits. (2) Invest-compete relation: In the robust regression, *competitive* is no longer significant, but business purchases is significant in the c-model. When clustering, *compete* fairly is significant in the a-and b-model, but not in the c-model. (3) Savings-knowledge relation: Robust and cluster regression leaves the significance level at 5-10% for *HIKknow*.

To get sensible log- results, we replaced zero values with the lowest of served value in the dataset. This was the case two times for sale, 17 for investments, three for purchases and seven for savings. Because of high numbers of negative values, we did not log profit.

4.5 Other analytical procedures considered

We considered and tried a range of methods for analysing the relationships we investigate in this paper. Firstly, we made ranked lists of all the participants ranking them by “entrepreneurality”. We wanted to compare this list to a ranked list of success based on the field data. As the data was quite small, and the dimensions so complicated, the results were not too interesting. As we have shown, the entrepreneurs faced so many additional constraints, so the analysis became non-conclusive.

Secondly, we wanted to use the panel data structure to run regressions and analyse the growth and dynamics of the data. The regressions did not prove to have much significance, in our small pool of participants, apart from the overall growth in the dataset, which we

decided to include in the main text for description purposes (see section 3.2.4). Additionally, the dataset is quite unbalanced, has gaps between the weeks of the year, and analysing weekly and daily data resulted not that interesting. In the SL Project the financial diaries were used as a technique of teaching the participants how to keep reports, not only for collecting data. Thus, the first entries of diaries data have somewhat lower quality than the later ones, which is another reason why we look at weekly *averages*.

Thirdly, to try to tie the lab indices and sub-indices as close as possible to behaviour in the field, we tried to find aspects of the financial diary reporting that could reflect behaviour directly. As an example, we tried to correlate knowledge against the standard deviation of the calculation mistakes in their financial diaries (they reported both detailed information and totals for many of the variables). This was however misleading, since some of the participants with lower calculation skills would sometimes leave the totals blank, and get lower standard deviations. Another example was correlating competitiveness to the number of entries since this was a criterion for winning the SL competition. The correlations did not provide any conclusions, and there were so many other reasons than a lack of competitive behaviour that kept them from handing in all the entries.

Finally therefore, we landed on the method used in this paper.

5. Summary and concluding remarks

Our paper aimed to provide insight into the lab-field relationship by investigating 49 young entrepreneurs involved in the Social Lab Project in the urban setting of Dar es Salaam. By comparing their lab performance in the previous Ruka Juu project with their field performance and choices, we provide suggestions for which entrepreneurial traits tend to be more important for outcomes and choices in a complex real world.

Lab experiments are widely used in the field of entrepreneurship. However, the literature on the link between lab experiments and field outcomes and choices is still remarkably limited. Our thesis is built on and inspired by the paper “*Competitive in the lab, successful in the field?*” by BBPT (2014), and our findings are in agreement. By applying the analysis on a different set of field data, we can verify their findings.

Our findings suggest that *competitiveness* is the key entrepreneurial trait for successful business outcome and entrepreneurial choices in the field. We find less consistent correlation with some of the other indices. The literature on competitiveness is generally focused on gender difference. In line with empirical studies, we observe a tendency of male participants outperforming the female participants. However, the differences are not significant.

Even though the effect on business outcome was weak, we find knowledge to be the most important entrepreneurial trait for savings.

As our data is not a representative sample for the common population, our findings cannot necessarily be generalized. However, due to similarity with the RJ dataset, the results from our data can still be useful for predictions in a sample comparable to the RJ survey pool.

Our paper cannot bring clear answers for policy changes, but may contribute in (1) narrowing the gap between experimental lab results and the outside world, and (2) giving indications to the current debate on entrepreneurship in developing countries on what skills and abilities of the young entrepreneur that are decisive for outcomes and choices. Our paper might give directions to future studies within the topic of entrepreneurship and microfinance, which has typically been concentrated around the aspects of financial and human capital.

References

- Almås, I., Cappelen, A. W., Salvanes, K. G., Sørensen, E. Ø., & Tungodden, B. (2014). Willingness to compete : family matters: Norwegian School of Economics. Department of Economics.
- Awinia, C. S. (2014). Structural Barriers, Constraints, and Urban Youth Employment The Case of Ilala Municipality, Dar-es-Salaam. Tanzania, : REOPA
- Backhaus, J. G., & Schumpeter, J. A. (2003). Joseph Alois Schumpeter: entrepreneurship, style, and visionEuropean heritage in economics and the social sciences (Vol. v. 1). Boston: Kluwer Academic Publishers.
- BBPT, Berge, L. I. O., Bjorvatn, K., Pires, A. J. G., & Tungodden, B. (2014). Competitive in the lab, successful in the field?
- Benz, M., & Meier, S. (2011). Do People Behave in Experiments as in the Field? - Evidence from Donations. *SSRN Working Paper Series*. doi: 10.2139/ssrn.756372
- Bjorvatn, K., Cappelen, A. W., Sekei, L. H., Sørensen, E. Ø., & Tungodden, B. (2013). *Edutainment for entrepreneurship: Evidence from a field experiment in Tanzania*.
- Bjorvatn, K., Falch, R., & Hernæs, U. (2012). Kvinner, Kontekst og Konkurransen - funn fra et labeksperiment i Uganda,. *Magma*.
- Bundervoet, T. (2013). Should We Stop Calling \$1.25 a Day Extreme Poverty? Retrieved from <http://blogs.worldbank.org/futuredevelopment/should-we-stop-calling-125-day-extreme-poverty>
- Buser, T., Niederle, M., & Oosterbeek, H. (2014). Gender, Competitiveness, and Career Choices. *The Quarterly Journal of Economics*, 129(3), 1409-1447. doi: 10.1093/qje/qju009
- Business Dictionary. (2014). Formal sector. Retrieved 09.12.2014, 2014, from <http://www.businessdictionary.com/definition/formal-sector.html>
- Central Intelligence Agency. (2014). The World Factbook
- Colwell, D., & Gillett, J. (1982). 66.49 Spearman versus Kendall. *The Mathematical Gazette*, 307-309.
- Cowling, M. (2007). Early stage survival and growth *The life cycle of entrepreneurial ventures* (pp. 479-506): Springer.
- Croson, R., & Gneezy, U. (2009). Gender Differences in Preferences. *Journal of Economic Literature*, 47(2), 448-474. doi: 10.1257/jel.47.2.448
- De La Paz Hernández Girón, J., León, M. Y., & Domínguez Hernández, M. L. (2007). FACTORES DE ÉXITO EN LOS NEGOCIOS DE ARTESANÍA EN MÉXICO. (Spanish). *Success factors of the handcraft business in Mexico. (English)*, 23(104), 77-99.
- de Mel, S., McKenzie, D., & Woodruff, C. (2008). Who are the microenterprise owners ? Evidence from Sri Lanka on Tokman v. de Soto: The World Bank.
- Deaton, A. (1989). Saving in developing countries: theory and review.
- Dhliwayo, S., & Van Vuuren, J. J. (2007). The strategic entrepreneurial thinking imperative. *Acta Commercii*, 7, 123-134.
- Dolgun, A. (2012). Practicum: Correlation. Hacettepe University, Faculty of Medicine Department of Biostatistics.
- Duflo, E., Banerjee, A., Glennerster, R., & Kinnan, C. G. (2013). The miracle of microfinance? Evidence from a randomized evaluation: National Bureau of Economic Research.
- Falk, A., & Fehr, E. (2003). Why labour market experiments? *Labour Economics*, 10(4), 399-406. doi: [http://dx.doi.org/10.1016/S0927-5371\(03\)00050-2](http://dx.doi.org/10.1016/S0927-5371(03)00050-2)

- Gerald, K., & Keller, G. (2012). *Managerial statistics* (9th ed. ed.). Australia: South-Western.
- Gneezy, U., Niederle, M., & Rustichini, A. (2003). PERFORMANCE IN COMPETITIVE ENVIRONMENTS: GENDER DIFFERENCES. *Quarterly Journal of Economics*, 118(3), 1049-1074. doi: 10.1162/00335530360698496
- Haber, S., & Reichel, A. (2005). Identifying Performance Measures of Small Ventures—The Case of the Tourism Industry. *Journal of Small Business Management*, 43(3), 257-286. doi: 10.1111/j.1540-627X.2005.00137.x
- ILO, I. L. O. (2013). Global Youth Employment Trends 2013- a generation at risk. Geneva.
- Lazear, E. P. (2005). Entrepreneurship. *Journal of Labor Economics*, 23(4), 649-680.
- Leppel, K., Williams, M., & Waldauer, C. (2001). The Impact of Parental Occupation and Socioeconomic Status on Choice of College Major. *Journal of Family and Economic Issues*, 22(4), 373-394. doi: 10.1023/A:1012716828901
- Levitt, S. D., & List, J. A. (2007). What Do Laboratory Experiments Measuring Social Preferences Reveal About the Real World? *Journal of Economic Perspectives*, 21(2), 153-174.
- Liedholm, C. (2002). Small Firm Dynamics: Evidence from Africa and Latin America. *Small Business Economics*, 18(1-3), 227.
- Loayza, N., Schmidt-Hebbel, K., & Servén, L. (2000). Saving in Developing Countries: An Overview. *The World Bank Economic Review*, 14(3), 393-414. doi: 10.2307/3990078
- Lærd Statistics. (2014). Spearman's correlation using Stata. *Stata Tutorials*. Retrieved 04 December, 2014, from <https://statistics.laerd.com/stata-tutorials/spearman-correlation-using-stata.php>
- Mensah, J. V., Tribe, M., & Weiss, J. (2007). The small-scale manufacturing sector in Ghana: a source of dynamism or of subsistence income? *Journal of International Development*, 19(2), 253-273. doi: 10.1002/jid.1322
- Mgumia, J. (2014). *Young People and Entrepreneurship in Tanzania: From Randomized Control Trial to Ethnography*. Working Paper. Norwegian School of Economics and University of South Africa.
- Ministry of Education and Vocational Training. (2012). *Tanzania education sector analysis. Beyond Primary Education, the Quest for Balanced and Efficient Policy Choices for Human Development and Economic Growth Tanzania*
- Neneh, N. B. (2012). An exploratory study on entrepreneurial mindset in the small and medium enterprise (SME) sector: A South African perspective on fostering small and medium enterprise (SME) success. *African Journal of Business Management*, 6(9), 3364-3372.
- Niederle, M., & Vesterlund, L. (2007). Do Women Shy Away From Competition? Do Men Compete Too Much? *The Quarterly Journal of Economics*, 122(3), 1067-1101. doi: 10.1162/qjec.122.3.1067
- Paige, R. C., & Littrell, M. A. (2002). Craft Retailers' Criteria for Success and Associated Business Strategies. *Journal of Small Business Management*, 40(4), 314-331.
- Ravallion, M., Chen, S., & Sangraula, P. (2009). Dollar a Day Revisited. *World Bank Economic Review*, 23(2), 163-184. doi: 10.1093/wber/lhp007
- Stata. (2014). pcorr — Partial and semipartial correlation coefficients. stata.com: Stata Data Analysis and Statistical Software.
- Statistics Solutions. (2014). Kendall's Tau and Spearman's Rank Correlation Coefficient. *Non-Parametric Analysis*. Retrieved 05 December, 2014, from

- <http://www.statisticssolutions.com/kendalls-tau-and-spearman-rank-correlation-coefficient/>
- Sutter, M., Kocher, M. G., Gltzle-Rtzler, D., & Trautmann, S. T. (2013). Impatience and Uncertainty: Experimental Decisions Predict Adolescents' Field Behavior. *American Economic Review*, 103(1), 510-531. doi: 10.1257/aer.103.1.510
- The National Examinations Council Of Tanzania NECTA. (2014). Examination types Retrieved 03.12 2014
- Toledo-López, A., Díaz-Pichardo, R., Jiménez-Castañeda, J. C., & Sánchez-Medina, P. S. (2012). Defining success in subsistence businesses. *Journal of Business Research*, 65(12), 1658-1664. doi: 10.1016/j.jbusres.2012.02.006
- UNDAP Tanzania. (2013). UNDP ANNUAL REPORT 2012-2013 *United Nations Development Assistance Plan (UNDAP)*.
- United Nations in Tanzania. (2014). Education Retrieved 02.12 2014 from <http://tz.one.un.org/index.php>
- Wooldridge, J. M. (2013). *Introductory econometrics: a modern approach* (5th ed., international ed. ed.). S.l.: South-Western, Cengage Learning.
- World Bank. (2014 -a). Harnessing Tanzania's Rapid Urbanization to Create Jobs, Thriving Cities. Retrieved 03.12, 2014, from <http://www.worldbank.org/en/country/tanzania/publication/tanzania-economic-update-harnessing-rapid-urbanization-create-jobs-thriving-cities>
- World Bank. (2014 -b). Tanzania Economic Update. Who Wants a Job? The Magnetic Power of Cities (Vol. 5): African Region Poverty Reduction & Economic Management
- World Bank. (2014 -c). Tanzania Overview Retrieved 02.12 2014 from <http://www.worldbank.org/en/country/tanzania/overview - 1>
- XE. (2014). XE (Historic) Currency Table: TZS - Tanzanian Shilling. Retrieved 12 December 2014, 2014, from <http://www.xe.com/currencytables/?from=TZS&date=2013-01-01>
- Zhang, J. Y. (2013). Can Experimental Economics Explain Competitive Behavior Outside the Lab? *SSRN Working Paper Series*. doi: 10.2139/ssrn.2292929

Appendix

Table of contents - Appendix

List of tables - Appendix.....	69
A. Normality test.....	70
A.1 Q-plots for normality checking	71
B. Financial Diary questionnaire.....	72
C. List of participants, business and lab performance	76
D. Weekly business development over the year	78
E. Supplementary correlation tests on competitive	79
E.1 Partial correlation on competitive	79
E.2 Point biserial correlation test on competitive	80
F. Regression results	81
References	83

List of tables - Appendix

<i>Table A-1: Skewness/Kurtosis test along with Shapiro-Wilk test for normality</i>	70
<i>Table C-1: The participant's score on the lab indices and type of business</i>	76
<i>Table E-1: Point biserial correlations test on competitive</i>	80
<i>Table F-1: Regression results of field outcomes on indices. All 49 participants</i>	81
<i>Table F-2: Regression results of field choices on indices. All 49 participants</i>	81
<i>Table F-3: Regression results of field choices on indices. Only the 29</i>	82
<i>Table F-4: Regression results field choices on indices. Only the 29</i>	82

A. Normality test

Below are the results of a skewness/kurtosis test of normality, and the Shapiro-Wilk normality test. The null hypothesis of the Shapiro-Wilk test is that the data tested is normally distributed. We therefore reject that there is evidence that the data is normally distributed at a low p-value. At a high p-value, one cannot draw any conclusions about the underlying distribution (JMP-StatisticalDiscovery, 2009). A Q-Q plot is required for verification of the test.

Table A-1: Skewness/Kurtosis test along with Shapiro-Wilk test for normality

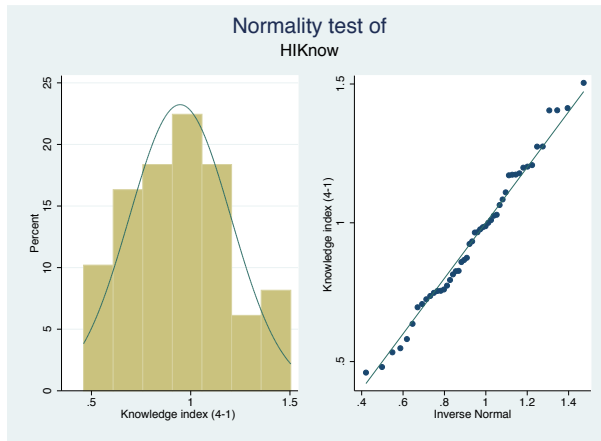
Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	Joint S an K test		Shapiro Wilk test
				adj chi2(2)	Prob>chi2	Prob>z
Age	28	0.2146	0.2508	3.15	0.2070	0.83428
Eat_meat	29	0.8588	0.6136	0.29	0.8665	0.99997
HIknow	29	0.1216	0.8820	2.65	0.2662	0.13756
know	29	0.5877	0.4245	0.99	0.6093	0.93020
mathp	27	0.2628	0.5918	1.68	0.4327	0.28916
cadd	29	0.1018	0.1040	5.18	0.0751	0.45624
entp	29	0.5566	0.9893	0.35	0.8388	0.99988
HIMinds	29	0.6899	0.9622	0.16	0.9224	0.98886
Patient	29	0.5920	0.7097	0.44	0.8035	0.99865
Risky	29	0.2789	0.6075	1.55	0.4607	0.99582
Confident	29	0.1137	0.3153	3.80	0.1495	0.06658
HIAmb	29	0.9696	0.1382	2.40	0.3010	0.99763
rank	29	0.9499	0.0126	5.85	0.0538	0.37405
mentor	29	0.8183	0.0201	5.28	0.0713	0.99038
Sales	29	0.0000	0.0000	42.53	0.0000	0.00000
Profits	29	0.0000	0.0000	26.87	0.0000	0.00000
Investments	29	0.0000	0.0005	22.26	0.0000	0.00000
Purchases	29	0.0000	0.0000	41.75	0.0000	0.00000
Savings	29	0.0000	0.0000	28.01	0.0000	0.00000

*Indicator variables, like *competitive*, have been taken out of the test, since it is clearly not normally distributed.

We see that many of the variables we use are rejected being normally distributed on a 5%significance level. All of our key field variables are rejected in both tests.

The tests find that we cannot reject that *HIknow* and the sub-indices *know*, *mathp*, and *entp* are normally distributed. The *cadd* (calculations) variable and the sub-variables of the *HIAmb* index are rejected based on a kurtosis-skewness test, but not in the Shapiro-Wilk test. We also tested for normality in our key field variables when converted to other functional forms, like log-form, inverse, quadratic and square. This does not help in achieving normality in the data.

A.1 Q-plots for normality checking

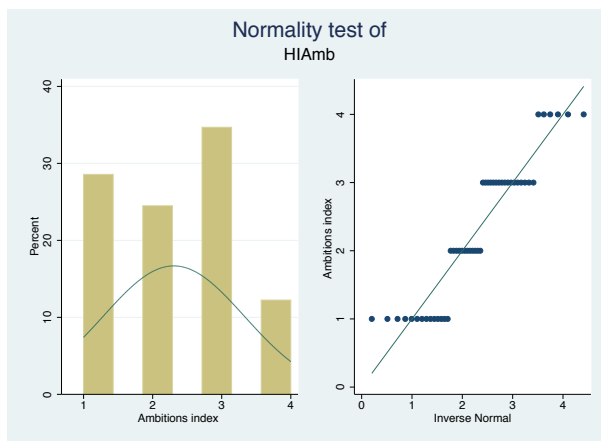


Now we will look more into the underlying distributions of some of the variables that are most important for our continued analysis.

HIKknow was not rejected in the normality test. *HIAmb* was rejected in one test, but not in the other. Graphing the distribution may help in understanding why. The graphs

below compare the distribution of the variable in question to the normal distribution. The histogram is supposed to follow the blue line, and the quantile plots are supposed to lie on

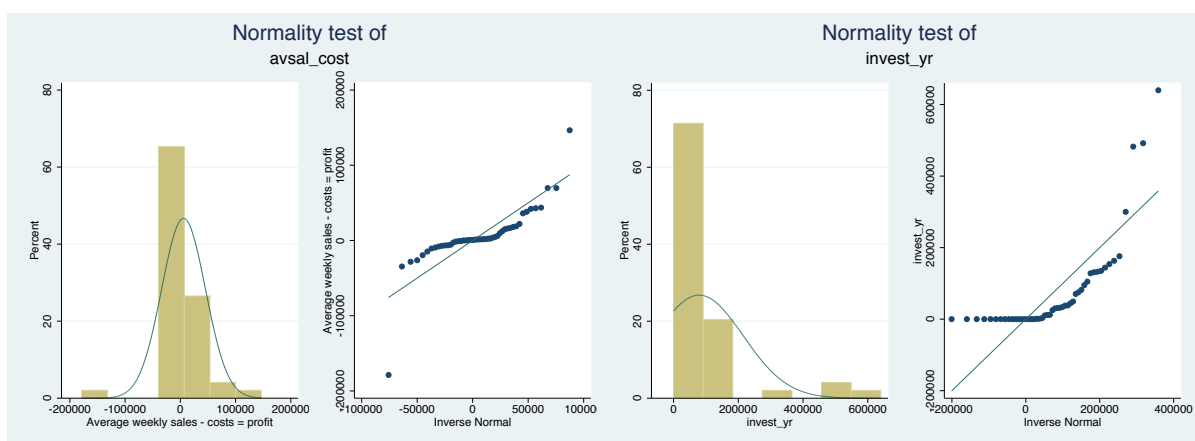
the line. This is a way of verifying the normality tests and to get an idea of the underlying distribution. The graph of *HIKknow* does not dramatically deviate from the normality distribution.



In the histogram on *HIAmb* we see that the distribution is somewhat skewed, but still quite peaked (kurtosis). Looking at this, it

does not seem that normality is an appropriate assumption for *HIAmb*.

Profits and investments seem to deviate from the normal distribution. The histogram on investments is very skewed, and both the Q plots deviate far from the line.



B. Financial Diary questionnaire

The Monthly Diary of:..... Week Number

Date.....

INCOMES AND EXPENDITURES

Table 1 (a): Daily Personal Income

	Mday	Tue	Wesn	Thurs	Fri	Sat	S'day
Business							
Parents							
Relatives							
Friends							
Jobs/Salaries							
Project							
Other: Specify							
Total							

Table 2 (a): Daily Personal Expenditures

	Mday	Tue	Wesn	Thurs	Fri	Sat	S'day
Food							
Drinks							
Transportation							
Phone/Communication							
Clothing and Footwear							
Fun/Entertainment							
Health							
Saloon							
Trips/Travels							
Alcohol/Cigarette							
Other							
Total							

Table 2 (b) Community/Family Expenditures

	Mday	Tue	Wesn	Thurs	Fri	Sat	S'day
Home: Specify							
1.1							
1.2							
Society (Social Events)							
2.1							
2.2							
Faith (Religion)							
3.1							
3.2							
Total							

BUSINESS: INVESTMENT COSTS, RUNNING COSTS, SALES, PROFITS

Table 3 (a) Investment Cost

	Narrative	Cost
1. Equipment investment- cooking equipment, carriages, machines etc		
2. Site (Premise or business site) – furniture such as selling table, renovation of space		
3. Tax /Public fees		
4. Other: Specify		

Table 3 (b) Running Cost

	Mday	Tue	Wesn	Thurs	Fri	Sat	S'day
Purchases (raw materials and goods)							
Transit of goods							
Commuting of person to/from business							
Communication							
Food and Drinks							
Rent							
Electricity							
Water							
Security							
Wages							
Tax							
Other							
Total							

Table 3 (c) Daily Sales

	Mday	Tue	Wesn	Thurs	Fri	Sat	S'day
Daily Sales							

Total sales per week.....

Table 3 (d) Profit and Stock Counts

State your profit per week	
State your stock per week	
State your loss per week	

Table 4 (Q 4.11 – Q4.27). YOUR FINANCIAL POSITIONS AND PRACTICES

What are your loans and savings balances at the end of the week?	Date	Loans (put date and name of service)	Savings (Put date and name of service)
Microfinance institutions			
Informal group Merry-go-round (Upatu)			
Formal credit and savings institution			
Bank			
Relatives or business partners			
Mobile			
Other (please specify, such as savings at home)			

TIME ORGANISING AND ACTIVITIES

Table 4 (Q 4 (M) – Q 4 (SU)): Organizing time (write number of hours)

	Mday	Tue	Wesn	Thurs	Fri	Sat	S'da y
Business							
Family							
Relatives							
Friends							
Social events							
Project activities							
Education							
Religion							
Leisure							
Health							
Sports							
Others: Specify							

4 (B) Please explain about the activity which you spend more time this week and why:

-Primary activity

-Reason

4 (C) Please explain which activity you liked most this week and why, and the one you liked least and why.

4Ci: Liked Activity:

- Reason:

4Cii: Disliked Activity:

- Reason

4 (d) Plans for next week: What are your activities plans for next week?

i)

ii)

iii)

5. (A) Reactions and Experiences

Table 5 (a) Scale: 1-5, where 1=very unhappy, 2=somewhat unhappy, 3=neither happy or unhappy, 4=somewhat happy, 5=very happy (if you choice 1 or 5 explain)

How happy are you with your family this months? (1-5)	
How happy are you with your economic situation? (1-5)	
How happy are you with your house economy? (1-5)	
How happy are you with your business/job (write N/A if person does not have income generating activity)	

5 (B) Mention challenges you faced this week?

i)

ii)

5 (c) What was your best/high-moment for this week?

i)

ii)

6) What are general plans that you are following up?

i)

ii)

C. List of participants, business and lab performance

Table C-1: The participant's score on the lab indices and type of business

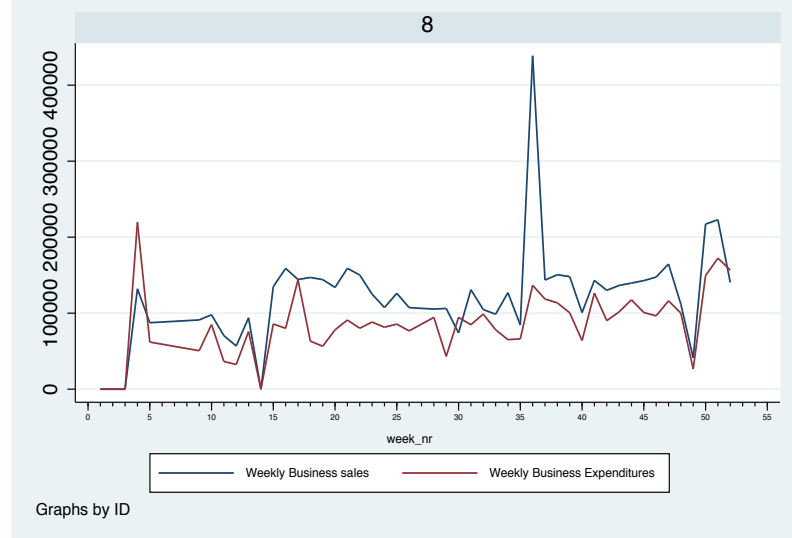
ID	HIKnow	HIMinds	HIAmb	Business
Participant 1	.7363828	1	3	Service (vegetable stand)
Participant 2	.8660641	2	3	Service (butcher)
Participant 3	.9764025	3	1	Service (shop)
Participant 4	1.173268	0	2	Service (DVD kiosk)
Participant 5	.9878491	4	1	Service (vegetable stand)
Participant 6	.8141888	2	1	Service (vegetable stand)
Participant 7	.755129	2	2	Service (water supply)
Participant 8	.7073296	1	3	Service (vegetable stand)
Participant 9	.5811385	2	4	Service (juice)
Participant 10	1.084146	1	3	Selling juice, fruits, changed to shoes
Participant 11	.7736589	3	1	Transportation (owns a boat) and kiosk
Participant 12	1.413271	2	3	Pregnant, never started
Participant 13	.7244582	2	2	Service (clothing shop)
Participant 14	1.50421	3	2	Selling clothes on credit
Participant 15	1.274493	1	1	Selling clothes on credit
Participant 16	.4804218	2	3	Fishing business
Participant 17	1.405425	0	3	Service (raising chickens, selling grains, running a café)
Participant 18	1.207833	2	2	Service (sell clothes on loan by distributing it to their network)
Participant 19	.9649974	0	1	Vegetable stand
Participant 20	.87397	1	3	Training centre
Participant 21	.9840989	2	1	Started selling nuts, but changed to the movie business.
Participant 22	.7543662	3	1	Clothes on credit
Participant 23	1.028998	3	2	Water business,
Participant 25	.7477608	2	2	Service (shoes and clothing business)
Participant 26	.6957301	2	1	Clothes on credits
Participant 27	.826968	2	2	Vegetable stand
Participant 28	1.274903	2	2	Service (raising chickens and selling them)
Participant 29	.7598245	1	3	Service (vegetable stand)
Participant 30	1.173721	1	2	Clothes on credit (T&D)
Participant 31	1.025533	2	4	Video shop
Participant 32	.4603789	2	1	Charcole then soap
Participant 33	.8255737	1	2	Service (selling DVD's)
Participant 34	.9660151	3	4	Pig business
Participant 35	1.171148	1	3	Vegetable stand
Participant 36	.5333089	1	1	Service (showcase football)

Participant 37	1.109722	2	3	Vegetable business at the market
Participant 38	1.202474	2	3	Lending out clothes
Participant 39	.6360697	2	3	Service (selling clothes at market place)
Participant 40	1.06406	0	4	Service (selling clothes)
Participant 41	.548107	3	1	Restaurant business
Participant 42	.7939446	3	3	Service (chips stand)
Participant 43	.9330135	1	3	Service (tailor)
Participant 44	.9237567	2	4	Textile (tie and dye)
Participant 46	1.198694	2	3	Service (raising chickens)
Participant 47	1.009741	0	2	Service (vegetable stand)
Participant 48	.9998924	1	1	Seafood, then a second hand business
Participant 49	1.404914	4	4	Service (bike transportation and supplying eggs)

D. Weekly business development over the year

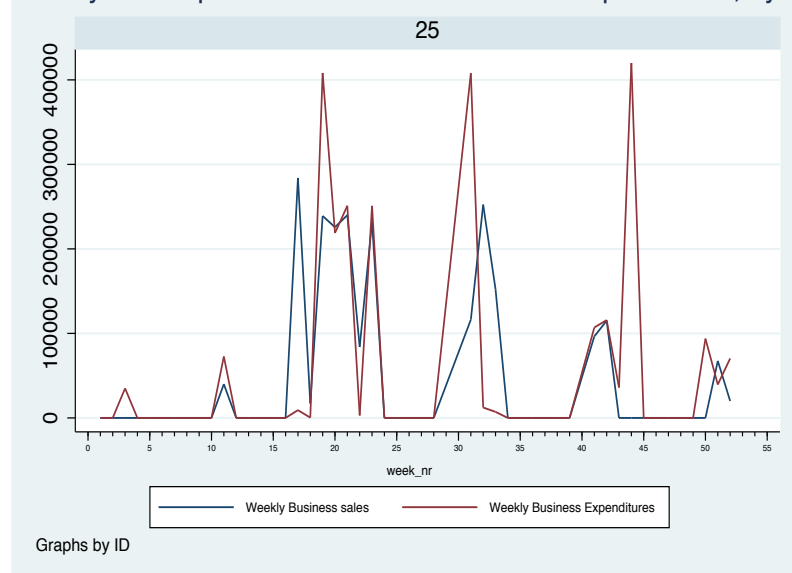
Below are two examples from two different sectors on weekly development on business sales and expenditures on individual level over the SLP year 2013.

Weekly Development of Business Sales and Expenditures, by id



Comment: The graph shows weekly development over the year for id 8, who run a vegetable stand. The peak observed in sales in week 37 is due to additional sales of artisanal products that he brought from his home village.

Weekly Development of Business Sales and Expenditures, by id



Comment: The graph shows weekly development for id 25, who sells clothes on credits.

E. Supplementary correlation tests on competitive

E.1 Partial correlation on competitive

For robustness, we will run some more tests on *competitive* to see if it stays strong.

As *competitive* was measured by asking if the participants wanted to *competitive* in mathematical calculations, one would assume that both mathematical skills (math grade and calculation skills) and confidence in how well they had done on the calculations compared to the others (*confident*), would affect this decision. “Willingness to take risk” is also likely to have an impact on this choice. We therefore ran partial correlations controlling for these variables on average business sales and yearly investment. Partial correlation measures the degree of association between two variables, and has the additional feature of removing the effects of other variables (Stata, 2014).

First, we run *competitive* against average sales, and control for each of the variables math grade, confidence and willingness to take risk at a time. *Competitive* is significant on a 5% level on all tests, except when controlling for math grade, where it is slightly over at p-value 0.0532 for business sales. When controlling for all variables at the same time, it has a significant correlation to business sales of 0.58. The correlation coefficient varies from 0.38 to 0.424.

When run against investments, *competitive* is still strong and significant on a 5% level when controlling for one variable at the time, and the correlation coefficient range from 0.38 to 0.42. When controlling for all variables, the correlation coefficient is 0.36, but is less significant on a 10% level.

Additionally, when we run Spearman correlations of *competitive* directly on these four variables for the whole dataset. We only find *confident* to have a correlation with *competitive* at a coefficient over 0.2. Confidence thus seems more influential when choosing to *competitive*, than actual mathematical skills.

E.2 Point biserial correlation test on competitive

Since our variable on *competitive* takes a value of 0-1, we decided to run a point biserial correlation to support our claims about the relation between this variable and the key field variables. This is a useful test when we want to correlate a binary variable with a ratio variable (Dolgun, 2012) like in our case. It uses a true Pearson product-moment correlation (Andersen, 1994).

We see that we still have strong correlations on average business sales, and it is almost significant at a 5% level. The correlation against investments is significant and even stronger, and against purchases strong, and significant at a 10% level.

Table E-1: Point biserial correlations test on competitive

Competitive on:	Coefficient:	P-val:	Obs.
-Sales	0.36	0.055	29
-Investments	0.44	0.002	49
-Business purchases	0.33	0.080	29

F. Regression results

Table F-1: Regression results of field outcomes on indices. All 49 participants

	Bus.survival		Sales		Profits			
			(1a)	(1b)	(1c)	(2a)	(2b)	(2c)
HIKnow		-0.192 (0.315)	-0.264 (0.330)	-16.736 (64.718)	-28.237 (71.017)		-3.235 (22.329)	-18.364 (23.184)
Patient	0.007 (0.056)	-0.011 (0.059)	-0.010 (0.057)	-11.700 (12.159)	-9.363 (12.510)	3.884 (12.876)	6.337 (4.292)	7.492 [*] (3.993)
Competitive	0.118 (0.173)	0.141 (0.179)	0.066 (0.177)	19.015 (37.255)	7.386 (38.736)	1.452 (39.185)	27.363 ^{**} (13.164)	19.837 (12.709)
Risky	-0.013 (0.107)	-0.004 (0.109)	-0.033 (0.108)	22.932 (22.336)	19.333 (22.683)	5.513 (24.152)	-0.284 (8.146)	-3.225 (7.722)
Confident	0.012 (0.032)	0.020 (0.033)	0.018 (0.033)	9.506 (6.722)	9.365 (6.974)	7.822 (7.241)	-3.216 (2.429)	-3.728 (2.354)
HIAmb		0.092 (0.079)	0.070 (0.077)	-18.429 (16.203)	-14.676 (16.605)		-15.388 ^{***} (5.568)	-15.468 ^{***} (5.423)
Female			-0.260 [*] (0.153)		-42.439 (34.527)			-10.755 (10.760)
Business_stream			0.283 (0.172)		-30.568 (38.153)			17.894 (12.097)
Newspaper			0.087 (0.156)		11.268 (34.346)			4.730 (10.951)
Own_bus			0.019 (0.153)		52.644 (34.149)			22.068 ^{**} (10.782)
Eat_meat			0.114 (0.080)		2.613 (18.190)			-10.218 [*] (5.646)
_cons	0.474 (0.340)	0.436 (0.462)	0.359 (0.527)	5.558 (69.936)	64.107 (94.817)	79.145 (113.902)	6.434 (25.903)	45.533 (32.706)
R ²	0.021	0.055	0.232	0.079	0.113	0.212	0.114	0.259
N	49	49	49	48	48	48	49	49

Std. errors in parenthesis

The regression on sales is excluding participant 2. The rest include all 49 participants.

Sales and profits are measured in 1000 TZS

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table F-2: Regression results of field choices on indices. All 49 participants

	Log investments			Log purchases			Log savings		
	(3a)	(3b)	(3c)	(4a)	(4b)	(4c)	(5a)	(5b)	(5c)
Patient	0.43 [*] (0.23)	0.29 (0.23)	0.31 (0.24)	0.36 (0.23)	0.35 (0.24)	0.32 (0.25)	0.17 (0.28)	0.27 (0.27)	0.40 [*] (0.23)
Competitive	1.94 ^{***} (0.72)	2.13 ^{***} (0.70)	2.17 ^{***} (0.75)	1.25 [*] (0.72)	1.11 (0.75)	0.91 (0.79)	-0.52 (0.86)	-0.23 (0.83)	-0.12 (0.71)
Risky	-0.08 (0.44)	-0.01 (0.43)	-0.06 (0.46)	-0.22 (0.44)	-0.27 (0.45)	-0.36 (0.49)	-0.54 (0.53)	-0.43 (0.51)	-0.59 (0.44)
Confident	-0.26 [*] (0.13)	-0.20 (0.13)	-0.18 (0.14)	-0.00 (0.13)	0.03 (0.14)	0.01 (0.15)	-0.22 (0.16)	-0.33 ^{**} (0.15)	-0.26 [*] (0.13)
HIKnow		-1.55 (1.24)	-1.83 (1.40)		-1.28 (1.31)	-1.18 (1.48)		4.01 ^{***} (1.46)	2.21 (1.33)
HIAmb		0.74 ^{**} (0.31)	0.81 ^{**} (0.33)		-0.01 (0.33)	-0.01 (0.35)		-0.26 (0.36)	-0.16 (0.31)
Female			-0.30 (0.65)			-0.04 (0.69)			-1.85 ^{***} (0.62)
Business_stream			-0.41 (0.73)			-0.38 (0.77)			1.23 [*] (0.69)
Newspaper			-0.65 (0.66)			0.61 (0.70)			-2.18 ^{***} (0.63)
Own_bus			0.26 (0.65)			0.50 (0.69)			0.79 (0.62)
Eat_meat			-0.09 (0.34)			0.32 (0.36)			-0.53 (0.32)
_cons	2.38 [*] (1.41)	2.08 (1.81)	3.32 (2.24)	1.40 (1.41)	2.60 (1.92)	1.39 (2.36)	4.29 ^{**} (1.69)	1.20 (2.14)	6.31 ^{***} (2.13)
R ²	0.211	0.314	0.342	0.122	0.142	0.188	0.071	0.212	0.516
N	49	49	49	49	49	49	49	49	49

Std. errors in parenthesis

Zero values have been given the lowest number in the sample. This applies for 17 for Investments, 3 for Purchases, and 7 for Savings

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table F-3: Regression results of field choices on indices. Only the 29

	Log sales (1a)	(1b)	(1c)
Patient	0.57* (0.30)	0.55* (0.31)	0.59 (0.34)
Competitive	2.51** (0.98)	2.68** (1.04)	2.77** (1.14)
Risky	-0.87 (0.66)	-1.02 (0.68)	-1.03 (0.73)
Confident	-0.33* (0.18)	-0.28 (0.18)	-0.28 (0.20)
HIKnow		-2.15 (1.81)	-2.42 (2.14)
HIAmb		0.06 (0.46)	0.17 (0.50)
Female			1.10 (0.97)
Business_stream			-0.65 (0.95)
Newspaper			-0.19 (0.99)
Own_bus			1.09 (0.94)
Eat_meat			-0.13 (0.46)
_cons	3.78** (1.79)	5.57** (2.60)	5.14 (3.23)
R ²	0.270	0.314	0.421
N	29	29	29

Std. errors in parenthesis

The regression on sales is excluding participant 2. The rest include all 29 participants.

Zero values have been given the lowest number in the sample.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table F-4: Regression results field choices on indices. Only the 29

	Log investments			Log purchases			Log savings		
	(3a)	(3b)	(3c)	(4a)	(4b)	(4c)	(5a)	(5b)	(5c)
Patient	0.35 (0.28)	0.26 (0.27)	0.36 (0.27)	0.67** (0.27)	0.61** (0.28)	0.67** (0.29)	0.40 (0.34)	0.48 (0.31)	0.62** (0.29)
Competitive	2.06** (0.92)	2.57** (0.91)	3.23*** (0.92)	2.27** (0.88)	2.63*** (0.92)	2.93*** (0.96)	1.09 (1.09)	0.52 (1.03)	0.79 (0.97)
Risky	-0.61 (0.61)	-0.65 (0.60)	-0.68 (0.58)	-0.87 (0.59)	-0.91 (0.60)	-0.83 (0.61)	-1.84** (0.73)	-1.58** (0.68)	-1.44** (0.62)
Confident	-0.26 (0.17)	-0.21 (0.16)	-0.21 (0.16)	-0.12 (0.16)	-0.08 (0.16)	-0.11 (0.17)	-0.41** (0.20)	-0.51** (0.18)	-0.48** (0.17)
HIKnow		-1.58 (1.59)	-2.35 (1.72)		-1.19 (1.60)	-0.68 (1.80)		4.41** (1.80)	3.16* (1.81)
HIAmb		0.77* (0.40)	1.01** (0.40)		0.54 (0.40)	0.52 (0.42)		-0.54 (0.46)	-0.26 (0.43)
Female			0.62 (0.78)			1.77** (0.82)			-0.78 (0.82)
Business_stream			-0.70 (0.77)			-0.18 (0.80)			0.64 (0.81)
Newspaper			-1.80** (0.80)			0.27 (0.84)			-1.86** (0.84)
Own_bus			-0.40 (0.76)			0.03 (0.79)			0.61 (0.80)
Eat_meat			0.02 (0.37)			0.03 (0.38)			-0.82** (0.39)
_cons	3.87** (1.67)	3.38 (2.29)	5.65** (2.60)	1.77 (1.61)	1.51 (2.30)	-0.43 (2.72)	6.35*** (1.98)	3.71 (2.59)	7.35** (2.74)
R ²	0.200	0.329	0.528	0.343	0.399	0.542	0.241	0.419	0.646
N	29	29	29	29	29	29	29	29	29

Std. errors in parenthesis

Zero values have been given the lowest number in the sample.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

References - Appendix

- Andersen, J. A. (1994). Point biserial correlation. *Stata Technical Bulletin*, 3(17).
- Dolgun, A. (2012). Practicum: Correlation. Hacettepe University, Faculty of Medicine Department of Biostatistics.
- JMP-StatisticalDiscovery. (2009). Usage Note 35406: How do I interpret the Shapiro-Wilk test for normality in JMP? *Support*. Retrieved 15 December, 2014, from <http://www.jmp.com/support/notes/35/406.html>
- Stata. (2014). pcorr — Partial and semipartial correlation coefficients. stata.com: Stata Data Analysis and Statistical Software.